https://doi.org/10.2352/issn.2168-3204.2017.1.0.88


# Evaluating Perceived Capture Quality for the Digitization of Cultural Heritage Objects

*Susan P. Farnand; Rochester Institute of Technology; Rochester, NY; Franziska Frey; Harvard Library; Cambridge, MA*

## Abstract

*Perceptual experiments were used to evaluate the capture quality attainable of four digitization systems. The study results showed that two of these systems produced images of limited use as digital masters. The perceived image quality for the other two systems was comparable for digitization purposes. While a variety of system characteristics must be given careful consideration when identifying equipment to purchase, a system unable to attain the perceived quality needed for usable images is of little value no matter how inexpensive and ergonomic it may be. Also, image quality cannot be defined by number of pixels. In this study, the system producing the largest files was not well rated. Using perceptual experiments helped clarify the utility of digitization systems.*

## Introduction

The Harvard libraries house an enormous quantity of material, both analog and digital, that is of keen interest to researchers around the world. To make as much of this available to as many of the interested parties as possible, several years ago Harvard embarked on an effort to digitize vast numbers of books, maps, papers, and other artifacts. To do this most efficiently and effectively, producing content that the researchers can actually use, the large-scale digitization process must be well understood and streamlined. To accomplish this, Harvard personnel are on a regular basis testing a variety of imaging hardware and potential workflows. The aim of this project is to develop a perceptual experimental technique for evaluating digitization systems and use this methodology to compare the capture quality attainable of custom reproduction equipment relative to commercially available complete systems.

## Methodology

The four systems in this experiment, one assembled from individual pieces including a DSLR on a copy stand with LED light banks with a CCT of D5500, and three turn-key systems: an image scanner, a copy stand with an incorporated single light bar and overhead camera, and a V-shaped book cradle overhead lighting and two point-and-shoot cameras. Each piece of equipment was set up and operated according to the manufacturer's specifications and procedures.

Harvard personnel created images from the same set of originals on each device. These originals included image types important to library digitization programs including colored illustrations, photographs, maps, books with marginalia, manuscripts, moiré sensitive prints, typewritten text, and several paper varieties, Figure 1.

With the images from the different pieces of equipment, a perceptual study was conducted in the laboratory to investigate the perceived quality of the various reproductions. This work was conducted with soft-copy reproductions to limit the variability added by the printing process. The images were prepared by cropping an area of each original, using a 3x4 aspect ratio, that would provide as close to full resolution on Device 1 as was practical while still giving appropriate content for the observers to evaluate. This device was chosen because the file size for this device was the second smallest, which would minimize the resizing needed for the other three devices. Also, its full size images tended to be slightly blurred. It seemed reasonable to limit the effects that resizing might have for this device. The images were resized using bicubic interpolation. The final images were displayed at 1200x1600 pixels. The cropping and resizing were performed using Adobe Photoshop. Resizing effects were not visually detectable, except in the moiré sensitive images. For these images care was taken to ensure that the moiré effects visible in the full size images were comparable in the resized images.

The experiment followed an anchored scaling protocol in which observers were asked to rate three test images relative to an anchor image that was assigned a value of 100. The images were shown side by side on a calibrated EIZO CG240 display with a Native P and Wyble, 1998). The colorimetric differences between the displays were found to be, on average, smaller than the perceptibility threshold in images of about 2.5 $\Delta E_{ab}$ units (Stokes et al., 1992). Both viewing areas have neutral gray walls. The room lights were off at RIT and at a dim level at Harvard.

A total of 34 observers participated in these experiments, 19 from the RIT campus environment and 15 at Harvard. These comprised primarily of professionals experienced with viewing and evaluating art objects and images including art historians, librarians, communications faculty and students, museum studies faculty and students, curators, imaging scientists, printers, photographers, and conservators. The observers color vision was tested using the Ishihara Plate Test of pseudo-isochromatic plates prior to initiation of the test. One observer was found to have a color vision deficiency. His results were not included with those of the 34 observers analyzed. All observers provided Informed Consent.

***Figure 1:*** *The image sets used in the study. From top left: Color Illustrations 3, Color Illustrations 4, Moiré Susceptible 2, Moiré Susceptible 3, Moiré Susceptible 4, Marginalia 1, Marginalia 2, Photograph 3, Photograph 2, Photograph1, Small Details - Text, Map 4, Map 3, Map 2, Manuscript, Paper Varieties 2, Paper Varieties 3, Paper Varieties 4, Paper Varieties 6, Paper Varieties 7, Paper Varieties 8*


***Figure 2:*** *Experimental interface*

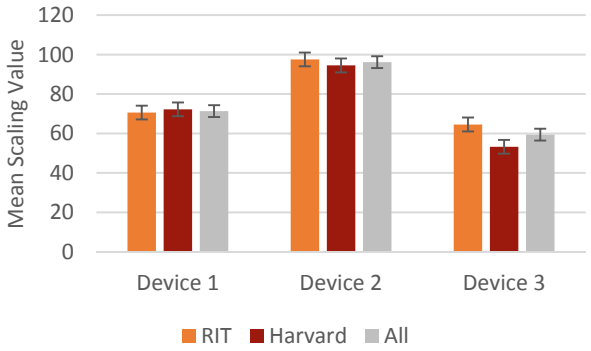


***Figure 3*** *The scaling values averaged over all 21 image sets for observers from RIT, Harvard, and overall*

## Results

The experimental results are summarized in Figure 3 and in Table I. These results indicate that the Anchor system, which was assigned the value of 100, and Device 2 were statistically significantly better than Devices 1 & 3 and that Device 1 was statistically significantly better than Device 3. The results also suggest that the Anchor was preferred over Device 2. This difference is less clear. It is possible that this is true. Several observers commented that they felt the images from Device 2 were too dark or were 'muddy'. However, the statistically significant difference between these two devices might not hold up if Device 2 were used as the Anchor. It is possible that observers, though instructed verbally and in writing, occasionally deferred to the Anchor as being a standard. In the author's opinion, further testing is needed to confirm this difference.

The data were evaluated to determine if there were differences in how the RIT and Harvard observers scaled the images. The results indicate that the differences were not significant for Devices 1 & 2, when averaged over the scenes. However, for Device 3, Harvard observers were statistically significantly harsher than RIT observers. This may be because the RIT observers included students, who had fewer years of experience working with historical documents.

**Table I: The maximally and minimally rated images for each of the digitization systems**

**Device 1**

| | | RIT | | Harvard |
|---|---|---|---|---|
| Max | Moiré Susceptible 4 | 106.4 | Moiré Susceptible 4 | 110.3 |
| Min | Text | 42.1 | Text | 44.7 |

**Device 2**

| | | RIT | | Harvard |
|---|---|---|---|---|
| Max | Moiré Susceptible 4 | 129.7 | Moiré Susceptible 4 | 122.7 |
| Min | Photograph 2 | 79.6 | Photograph 2 | 75.3 |

**Device 3**

| | | RIT | | Harvard |
|---|---|---|---|---|
| Max | Moiré Susceptible 4 | 105.8 | Paper Variety 2 | 93.3 |
| Min | Photograph 3 | 41.7 | Photograph 3 | 35.0 |

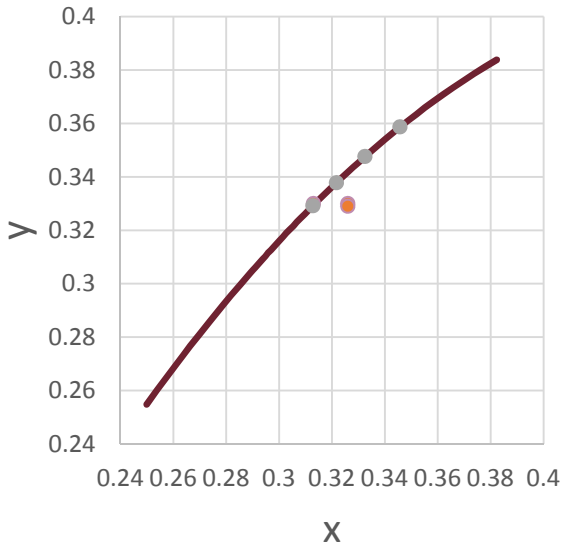


***Figure 4*** *Display white points of the Harvard and RIT displays (red) relative to the blackbody curve on the xy chromaticity diagram. Gray points depict daylight white point*s

Table I lists the maximum and minimum ratings that were assigned, on average, by the RIT and Harvard observers for each of the digitization systems. Almost all of the maximum rating values were assigned to the Moiré Susceptible 4 images. Because values greater than 100 indicate better performance than the anchor, this indicates that the Anchor device performed poorest for this image set and that, consequently images prone to moiré or aliasing might be difficult for this system to acceptably reproduce. The one digitization system and observer set that did not receive the maximum rating for the Moiré Susceptible 4 image set was Device 3 for the Harvard observers. In this case, the observers rated one of the Paper Varieties the highest. This paper was a type written sheet with a high degree of show-through. The images for Device 3 were typically rated below 65 because this system tended toward over-exposure of the originals, yielding a reproduction in which the text was clear, the paper background was bright, and the show-through was eliminated. Some observers preferred this reproduction approach for this particular original. The only two other images that rated above 65 were of moiré susceptible original. This device produced high resolution images that did not show the aliasing artifacts of the other systems. At the other end of the ratings scale, the image sets that received the lowest ratings differed for the three systems. For Device 1, the lowest rated image set was the Small Details Text image. This digitization system tended to produce blurry images. Observers found this unappealing, especially in the case of text. The ratings for Device 1 were almost all in the range of 64-84. The exceptions on the low side were all Small Details images. The two exceptions on the high side were Moiré susceptible images. For these sets, the blur masked some of the aliasing artifact, which led observers to rate this type of image higher.

Interestingly, two of the three lowest rated images for Device 2 were the other two moiré susceptible images. (The other 18 images were all rated above 85.) So, this digitization system also had difficulties with aliasing. However, the lowest rated image was one of the photographs. This photograph had a background area that Device 2 reproduced slightly pink, which observers did not favor. This pink cast may have resulted from the display calibration, which was slightly to the pink side of the blackbody curve as plotted on a chromaticity diagram, Figure 4. This calibration worked well for the other three systems, but yielded slight pink results for Device 2. If an institution chose to use this system, the default sRGB calibration may likely be used successfully.

The lowest rated image for Device 3 was also a photograph. This device, as already mentioned, tended toward overexposure. Its color and tone reproduction were referred to with terms like ‘jarring’, ‘washed out’, and ‘headache-inducing’. This trend was particularly unappealing in the reproduction of photographs. It was much more successful in the reproduction of text, especially if the observers felt that the digital master would be used to make copies for multiple generations – copies of copies of copies. For this application, this digitization would work well

Along with the scaling results, observers provided semantic data. The observers were asked to describe their scaling criteria for images from 9 of the 21 image sets, covering the range of the original types, specifically: Color Illustrations 3, Manuscript, Moiré Susceptible 2, Marginalia 1, Marginalia 2, Text, Photograph 1, Map 2, Paper Varieties 3, Paper Varieties 8. Most of their responses could

be categorized as falling into three areas: color, sharpness, and contrast, Table II. The sharpness category included words like blur, legibility, readability, and clarity. The contrast category included descriptions like exposure, tone, blown-out, and loss of shadow detail. Two other important categories involved the authentic or natural appearance and the ability to see the paper details. Many observers wanted the reproductions to accurately represent the original object and these two categories of description were important factors in this regard. The results in Table II show that sharpness was mentioned more than any other image quality attribute. Other interesting comments for each image type are included in Table III. One comment of particular interest was that observers described wanting to see the depth of the ink as well as the paper structure for the Paper Varieties 8 image, which the Anchor device did the best job of reproducing.

**Table II: Semantic results for the image types evaluated in this study in terms of the number of times observers used each image quality characteristic in their descriptions**

| | Illustration | Manuscripts | Marginalia | Moire Susceptible | Paper Varieties | Paper Varieties | Photos | Text | Map | Totals |
|---|---|---|---|---|---|---|---|---|---|---|
| Color | 27 | 8 | 8 | 27 | 6 | 4 | 7 | 9 | 18 | 114 |
| Sharpness | 19 | 29 | 28 | 7 | 26 | 22 | 15 | 25 | 29 | 200 |
| Contrast | 10 | 18 | 16 | 2 | 19 | 10 | 23 | 16 | 15 | 129 |
| Authentic | 5 | | 1 | 15 | | | 8 | 1 | | 30 |
| Show-thru | | | 6 | | 1 | | | 1 | | 8 |
| Distortion | | | 3 | | | | | | 2 | 5 |
| PaperDetails | | 11 | | | 6 | 24 | | | | 41 |
| Addenda | | | 12 | | 1 | | | | | 13 |

**Table III: Comments made in the semantic results for each image type evaluated in this study**

| | |
|---|---|
| **Illustration** | Color was important for this image type. Many wanted the color to look natural, though others preferred the more saturated color |
| **Manuscripts** | Multi-generations may be important; handwriting not as demanding as text |
| **Marginalia** | Some wanted to see the show thru while some felt it was distracting |
| **Moire Susceptible** | Some felt this could be too sharp - dots too distinct while some felt it was important to see the dot structure in this image type |
| **Paper Varieties 2** | Some wanted to see pencil and stamp details |
| **Paper Varieties 8** | Many wanted to see the details of the paper, some mentioned wanting to see depth of ink on the paper |
| **Photos** | Shadow detail mentioned at least 5 times, one mentioned thinking of what this reproduction would look like in a book |
| **Text** | Multi-generations may be important for this image type |
| **Map** | Many wanted to see map details such as towns and rivers and boundaries |

## Conclusion

There is more to the digitization equipment decision than image quality. Criteria important in selecting a digitization system include factors like space requirements, network and electrical connectivity, ease of set up, ergonomics, productivity, range of materials that can be digitized, image and color quality, and cost. [4] All of these must be given careful consideration when identifying a system to purchase. Clearly, however, a system with inadequate image quality, which will not provide usable digital masters, is of little value, no matter how inexpensive and easy to use it may be. One would not currently use a smartphone on a tripod as a digitization device. (Though, with the strides being made with these cameras, this may not be the case in the future.) In this study, it was found that one of the turn-key digitization systems produced blurred images that would not be usable as digital masters. Only in the case of moiré susceptible originals, where the blur smoothed out some of the patterning artifacts, was this system close to the performance of the anchor system. Even with these originals, however, observers objected to the loss of detail in the reproductions. Another of the commercially complete systems also did not reproduce originals adequately for use as digital masters. In this case, the images were greatly over-exposed. This system provided high resolution images with clear, sharp text, which might make it useful for capturing moiré susceptible originals, text originals, such as books, and originals where the digital masters may be the first of several generations of reproductions. This system, however, generally loses the sense of the original artifact including color, bleed-through, marginalia, light lines, and paper structure. In general, the perceived quality was low, despite its large image files, highlighting that there is more to image quality than pixels. The third system performed similarly to the custom reproduction equipment, which served as the anchor system, though it had some problems with plain text documents because it tended to produce images that were a little darker than the anchor system as a result of illumination limitations. And, both this system and the anchor had difficulty with moiré susceptible originals. However, overall, both of these two systems, in this experiment, produced results indicating that they would be reasonable to choose for digitization equipment.


## Acknowledgements

The authors thank would to thank Imaging Services and Preservation Services staff at the Harvard Library for assembling and imaging the originals and providing assistance in a myriad of other ways. We are also grateful for the time and the insights provided by all of our observers.


## References


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

[2] Fairchild, M.D. and D. Wyble, (1998) Colorimetric characterization of the Apple studio display (Flat panel LCD), http://scholarworks.rit.edu/article.

[3] Stokes, M., Fairchild, M. D., & Berns, R. S. (1992). Precision requirements for digital color reproduction. *ACM Transactions on Graphics (TOG)*, *11*(4), 406-422.

[4] Personal conversation with David Remington, Harvard Library.


## Author Biography

*Susan Farnand is a Visiting Assistant Professor in the Program of Color Science at the Rochester Institute of Technology. Her research interests include human vision and perception, color science, cultural heritage imaging and 3Dprinting. She received her BS in engineering from Cornell University, her Masters in Imaging Science and her PhD in Color Science from the Rochester Institute of Technology. She began her career at Eastman Kodak, designing and evaluating printer systems. She is Publications Vice President of the international Society of Imaging Science and Technology and serves as an Associate Editor for the Journal of Imaging Science and Technology. She participates in several Standards efforts including ISO TC 42 JWG26 Archival Imaging.*