Exploring the Use of Targets to Study Scanning Exposure Variability

F. Barry Wheeler; The Library of Congress; Washington, DC., USA

Abstract

Each year, the Library of Congress (Library) and its contractors create hundreds of thousands of digital images from documents held in the Library's collections. While a large variety of scanners and software are used by many different operators, the key determinant of image quality is proper exposure. In some form, exposure determines tonality, dynamic range, shadow and highlight detail, "sharpness," noise, and color accuracy. Accurate exposure is essential so it is somewhat surprising that so little attention has been paid to insuring common exposure calibration and standardization within and across the imaging workflows. This paper describes a new set of image targets with analytical software used to create and evaluate baseline scanner image quality. Then it begins examining the effects various exposure controls have on image quality.

Introduction

The Library of Congress and its contractors create hundreds of thousands of images each year as the documents and artifacts of our cultural heritage are imaged using a variety of scanners. Over time the Library has worked to achieve higher production levels while maintaining – and even improving – image quality. Many other federal agencies, state and local government institutions, and private organizations have developed similar conversion systems for the creation of new digital collections.

In an effort to continue improving image quality and productivity, the Library is working with almost a dozen government agencies that constitute the Federal Agencies Still Image Digitization Standards Working Group (Working Group). This working group is focused on developing standards, best practices, and tools to support imaging activities in the participating agencies. In connection with the work of this group, the Library is supporting the development of imaging targets designed to help achieve higher image quality and consistency. The targets include grayscale and color patches, with annotated CIELAB (LAB) values, ISO-defined slant-edge resolution and visual resolution (hyperbolic wedges along three axes) features, as well as dimensional scales on the object target. A software application accompanying the targets provides the detailed data and graphic displays necessary to determine and analyze the measurements supported by the various target elements.

Library of Congress Support of Imaging Standards

One initial activity of this Working Group is to review the imaging specifications developed by each agency. Imaging specifications - such as resolution, tone and color accuracy, noise levels - vary widely. Generally imagining project specifications depend upon the type of materials being digitized and the objectives of the project. Currently the Working Group is reviewing a draft overview relating content categories and digitization objectives to image-quality characteristics. An example is "Category 1. Textual and illustrated printed matter (books, journals, manuscripts, some maps) … Sub-category T.1. Clean, high-contrast book pages … [having] both informational and artifactual value…." This category is linked to a specific set of digitization objectives, expressed as use cases. Each use case is linked in turn to quality notes. These quality notes are designed to support the choice of appropriate imaging specifications and process controls monitored by the use of targets. Under this plan, the specifications for digitizing a plain text book from a general collection can be very different from the specifications for digitizing a rare book with hand-colored illustrations with each imaging operation monitored using targets.

In support of this work, the Library is developing components for an imaging toolkit incorporating "profiles" for customizing the tools to the specific project imaging specifications. The first tools under development are an automated image validator and a set of imaging targets with analytical software.

The Configurable Image Validator (CIV)

The Configurable Image Validator (CIV) is built on JHOVE, the JSTOR/Harvard Object Validation Environment software tool (available at http://hul.harvard.edu/jhove/). JHOVE analyzes and checks to ensure that images meet and maintain the requirements of a specific format. The Library has created a graphic user interface and profile builder utilizing the JHOVE engine. Each profile contains a list of the appropriate format, technical, and metadata specifications developed for the specific digitization project content and objectives. The CIV supports batch quality assurance inspection of every image file to ensure each file meets the required specifications.

Device, Object, and Microfilm Targets

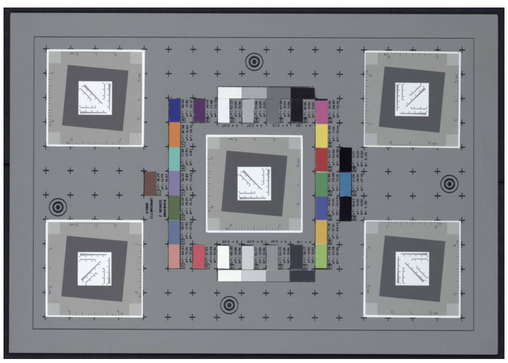
The Library is also supporting the development of reflective and transmissive targets for measurement of image resolution, noise, tonal characteristics, and color accuracy compliant with ISO and IEC imaging standards. The **Digital Image Conformance Evaluation (DICE)** system consists of **device** and **object** targets, (shown and described below) with software. Similar to the CIV profiles, performance profiles are built around these targets to store imaging specifications appropriate to the project content categories and objectives. The software analytical tools indicate if the equipment and workflow produce images that meet the requirements specified in the profile.

Targets, along with software analytical tools, provide objective measurements of scanner performance. Current ISO standards define measurements of resolution, tonality, dynamic range, noise, and color accuracy. The Library of Congress has engaged Don Williams to produce targets and software to support the Library's and other Federal agencies' imaging projects. Measurements obtained from the device target, shown in ***Fig. 1*** below, can provide definitive characterizations of scanner quality to guide purchase decisions and guide imaging operations. Individual measures derived from this target can ensure that scanner setup, configuration and operation produce the desired quality based on specifications developed from the content type and project objectives discussed earlier. An object target, shown in ***Fig. 2,*** can be included in each document scan. This target demonstrates continued scanner performance and can also provide the image recipient valuable data indicating that the image reflects the original document at the required specifications.

The Device Target

The device target used in this study was designed by Don Williams and Peter Burns. Peter Burns and Steve Tomanovich developed the software. This single target contains a series of features designed to measure resolution, noise, tonal values, dynamic range, color registration, color accuracy, and illumination uniformity. These technical features are accompanied by visual

Fig. 1 – The Device Target



indications of resolution and original document size. Noise, grayscale tonality, color values, color neutrality, artifacts and color fringing can also be judged visually.

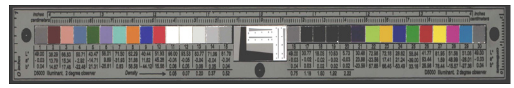
Software analytical tools accompanying the Library's targets apply current ISO standards to measure resolution as a *Spatial Frequency Response (SFR)* with complete curves and SFR(50) and SFR(10) values. The *Opto-Electronic Conversion Function (OECF)* curve is displayed showing white and black points, tonality, and dynamic range. The designers choose specific dark and light patches to help determine if detail is lost in those areas of the image. The grayscale patches also support measurement of noise and color neutrality. The color patches provide known colors with printed aim values and the software computes and graphs $\Delta E(2000)$ using L*a*b* values as well as separate computations and graphs based on a*b*.

The software also supports profiles that collect project specifications into a single configuration file. A profile allows specific aim values and acceptable limits to be entered. For instance, the user could enter 400 dpi plus or minus 10%. Or the user could stipulate 400 dpi minus 30 dpi / plus 100 dpi. This flexibility helps the scanner operator configure the equipment for imaging rare books using one profile and then re-configure the equipment for imaging a plain text manuscript. The software package provides a tabbed interface that simplifies creating profiles and reviewing image data using different profiles.

The Object Target

The object target includes the same features in a smaller package that can be included within the image capture field as documents are imaged. The use of the object target ensures that the required specifications are maintained throughout the production run.

Fig. 2 – The Object Target



Statement of the Problem

The primary objective of this project is to produce an image quality baseline of all overhead scanners used at the Library of Congress.

Currently the Library and its contractors operate at least nine overhead scanners. Ten additional high-production scanners are currently being installed, with additional high-production scanners possible. Materials are grouped and assigned to scanners according to current perceptions of image quality and production capability. Operations are managed based on experience gained over the past 10 to 15 years.

As evidenced by the rapid growth in high-production scanning projects, there is considerable interest in increasing image production while maintaining or improving image quality. There is also a desire to coordinate imaging projects with other Federal agencies and private operations such as the Internet Archive, Google Scanning and Microsoft Scanning. All operations are beginning to implement the recent ISO standards of image quality measurements. Scanner manufacturers are racing to improve the quality of their equipment to meet the new requirements and specifications.

The Library's new targets and software are designed to facilitate this progress in imaging. This paper presents an engineering effort to move the targets from the world of imaging science into the imaging production environment: a theory to practice transition. As such, it is attempting to achieve three objectives:

- To create a baseline of image quality -
- To facilitate matching scanners to projects on the basis of appropriate image quality for the document type and project objectives.
- To facilitate workflow and production operations to maintain the desired image quality over time.
- To provide for comparative analysis as imaging variables are managed to increase either quality or productivity.
- To provide a comparative analysis of software computed image quality measurements and human assessed image quality measures that will -
- Help refine the targets and software.
- Provide knowledge of and confidence in the targets and software.
- Support planning and developing the profiles used to evaluate equipment and production within the context of content category and project objectives.

- To demonstrate how to use the targets to optimize scanning variables, specifically focusing on how changes of lens aperture and ISO equivalent sensitivity (or digital gain) affect image quality.
- Lens aperture is often adjusted to maximize depth of field so that re-focusing is not required as frequently. Aperture adjustments also help determine white point placement and thus proper exposure.
- ISO-equivalent sensitivity is adjusted for white point placement and proper exposure.
- Obtaining a quality baseline and measures of how exposure variations affect image quality is a necessary first step to future analyses of how other variable affect quality – particularly post-processing actions such as contrast stretching or sharpening.

It is important to note that this project is not a "pure" research project. Variables cannot be cleanly isolated and controlled in the production environment. Most of the Library's scanning equipment incorporates all hardware, optical components and software in a single package that prohibits the control of some variables and limits the range of control of others. Rather than attempt a "laboratory" approach, current research in imaging science is applied within the ongoing Library production workflow and operations. It is hoped that this engineering process reaps the benefits of imaging science within the confines and structures of the Library's workplace.

Design

This study focuses on the use of the *device* target to collect baseline performance measures for each scanner under routine operational conditions. The ***planned*** data collection procedure followed this outline: **Setup -**

- Documentation of the scanner workspace including measurement of the environmental lighting, the document illumination, and any specific lighting to evaluate images. The scanner exposure settings, specifically for ISO sensitivity equivalent ("digital gain") and scanning speed, are noted. Additional scanner controls for resolution, tonal and color accuracy, sharpening and other image capture processing parameters are also examined.
- The scanner is set up for white point (exposure), color accuracy, and focus according to the manufacturer's instructions. The lens aperture is set according to standard operating procedures.
- The device target is placed on the scanner table.

To Collect Baseline Data:

- The device target is imaged at least once for each resolution at which the scanner is used.
- All routine standard operating procedures are followed.
- If these standard operating procedures include white point adjustment, the expectation is that the adjustment will be done by changing the *exposure time.*

Since this study is primarily intended to document exposure variability, the images collected in the next two steps have been limited to image sets at a single sampling rate, 400 dpi. This sampling rate was chosen because it is a common imaging resolution used on all scanners currently in production at the Library. Currently the Library specifies 400 dpi if OCR is to be performed on text included on the images, making it a common specification and thus the best resolution for this study.

For the Exposure Adjustment – Aperture:

- The scanner was set to 400 dpi with the lens at the aperture used for the baseline images. One or more images are captured for each variation below.
- The aperture was opened one stop – i.e., from f8 to f5.6 and the white point re-established, if possible by decreasing the exposure time.
- The aperture was opened an additional stop and again the white point re-established.
- The procedure was repeated by closing the aperture from the original setting in one stop increments.
- If the white point could not be re-established, the range of aperture changes was limited to a 1 stop variance without adjustment. (This provides an image suitable for examining the effects of post-processing although these effects are not included in this paper.)
- The aperture is reset to the value used for the baseline images.

For the Exposure Adjustment–ISO Sensitivity Equivalent:

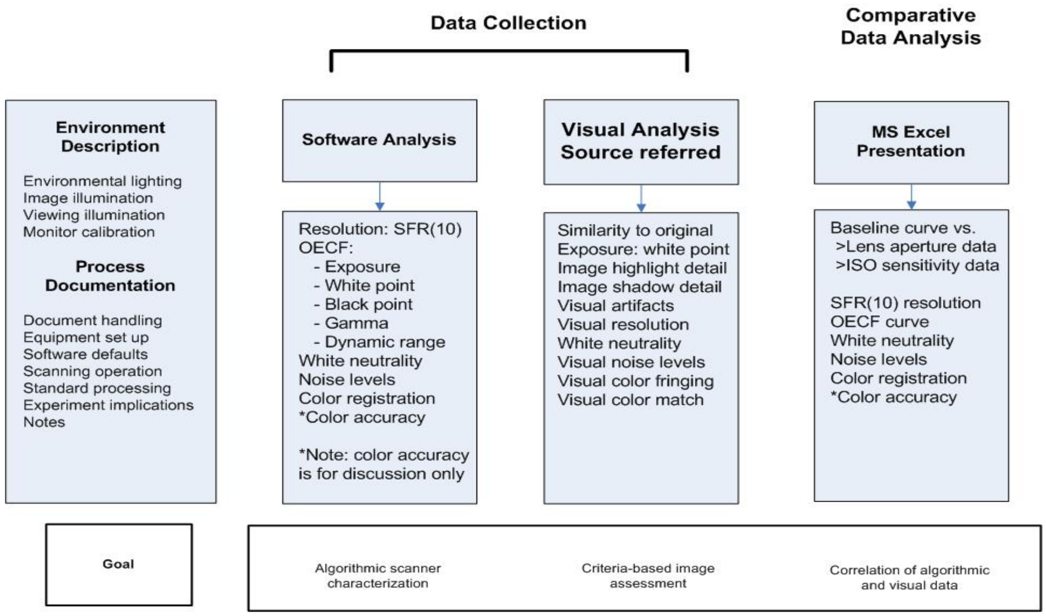
- The scanner was set to 400 dpi with the lens at the aperture used for the baseline images. One or more images is captured for each variation below.
- The ISO sensitivity was doubled (if possible), making the scanner twice as sensitive to light. The white point was re-adjusted, if possible by adjusting the exposure time. If the exposure time could not be adjusted, the aperture was closed one stop.
- The ISO sensitivity was re-doubled and the required adjustment made for white point.
- The ISO sensitivity was halved (if possible) from its original setting and the white point adjustment made by either increasing the exposure time or opening the aperture.
- After all images had been made the scanner was reset to the initial values used for the baseline images.

(The procedure is modified to accommodate the restrictions of the equipment. For example, for our *Jumbo Scan* a factory technician provided only two (2) configuration files for each standard resolution. Varying the aperture beyond one (1) stop produces completely unacceptable images and it is very hard to vary the ISO sensitivity equivalent. Thus only four (4) images were produced for this scanner.)

Data Collection

Fig. 3 summarizes the data collected and initial data analysis done for this study.

Fig 3. Examining Image Quality and Exposure Settings – Data Analysis



Environment

Environmental variables can frequently be controlled for a lab experiment. Not so in a production environment. Each of the Library's scanning operations is likely to have different room lighting, different scanner lights, different document viewing lights, and different monitor calibration standards. Hardware and software defaults and operating procedures differ as well. While process control is the primary goal, this is often unachievable in real world workflow environments. Consistency and performance documentation should be consider suitable secondary goals. Particularly important are the standard operating procedures (SOP) that govern scanner exposure and the software defaults that perform automatic processing during image capture.

All data collection begins with an initial scan following the routine procedures in use for that particular scanner. Aperture, exposure time, and ISO sensitivity variations begin from this SOP image. No special profiles are used in this study – only the actual data collection modules of the DICE software system are used.

Software reports

Fig. 3 lists the image parameters measured by the DICE software. Each measurement is based on the appropriate ISO standard; most measures require graphic presentation because they provide data across the complete tonal range and /or data within the complete RGB or LAB color models.

Visual reports

Data and charts are easier to comprehend when visual illustrations demonstrate the data points on real images. For this study, visual data representation is limited to elements shown on scans of the device target. (The DICE team is developing an image set illustrating different data elements on scans of real-world documents but real-world illustrations are beyond the scope of this report.)

An attempt has been made to pair each data presentation with a corresponding visual element or characteristic. Two such examples are included within the body of this report. A more complete presentation showing both software and visual data will be available at the conference or directly from the author.

Presentation

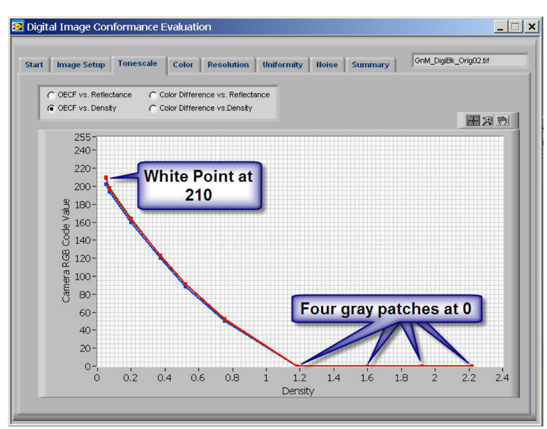
In addition to the graphical and numerical data presented within DICE, the software outputs an MS Excel formatted file of input data read from the image and calculations. (The actual calculations are performed by the National Instruments LabView 8.5 runtime engine used by the DICE software.) Thus a user can combine and analyze data and computed results comparing different images as well as present comparative results in tabular and graphical form.

Sample Results

Software analysis – OECF curve

Fig. 4 shows the *Opto-Electronic Conversion Function (OECF)* curve of a device target scan performed according to the SOP currently used for this machine. All manufacturer's procedures for setting focus, white point, and color balance were performed as usual for a 400 dpi image.

Fig. 4 – DICE Data Presentation, SOP Image

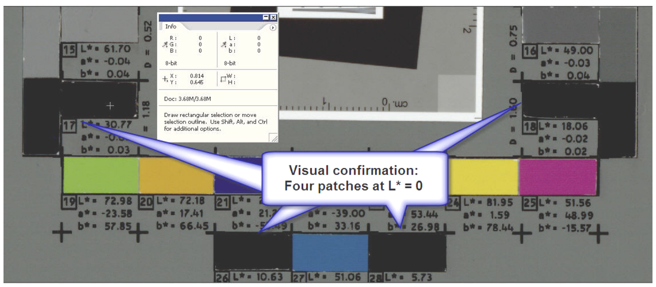


The L*a*b* luminance (L*) value of the white point on the device target is 96. Current Library standards would require the white point placement of this L* value between 243 and 250. The Luminance values for the four gray patches, from left to right, are L* = 30, 18, 10, and 5 respectively (rounded down to the nearest integer value).

Visual analysis – the OECF curve

Fig.5 shows the device target image area corresponding to the four dark patches. In the original print all four patches are black with no visible differentiation among the patches. Current Library standards would require that no detail be lost in the shadow (dark) areas. Given this target data it appears that dark area detail will be lost.

Figure 5 – Visual Confirmation of Fig. 2 Data



Note that these scans may be acceptable – even preferred in some cases. Content type and project objectives provide the basis for the imaging specifications. Books from the general collection might benefit from the tonal compression of black and dark gray showing clearer text. Maps may show boundaries and line detail with improved clarity because of these tonal curve characteristics – a welcome tonal adjustment for users trying to research the content of the map. The use of profiles allows the project manager to customize the image specifications based on Working Group schema for document type and project objectives. The data and graphs then present information on how the specifications impact image quality, including capture data loss. Above all, the goal for any collection digitization process is consistency. Preferences may differ, but precision must be maintained – data consistency makes data management tasks much easier.

Software analysis – SFR curves

Fig. 6 shows the SFR curves for the same image. Current Library standards set requirements for the SFR(10) level. The curves in *Fig.6* show that the luminance, blue and green channel graphs cross the SFR(10) – i.e., .01, level above the required 400 dpi specification while the red channel crosses the line below 400 dpi. The average dpi is shown in the upper right corner as 399.08 dpi – definitely within acceptable limits.

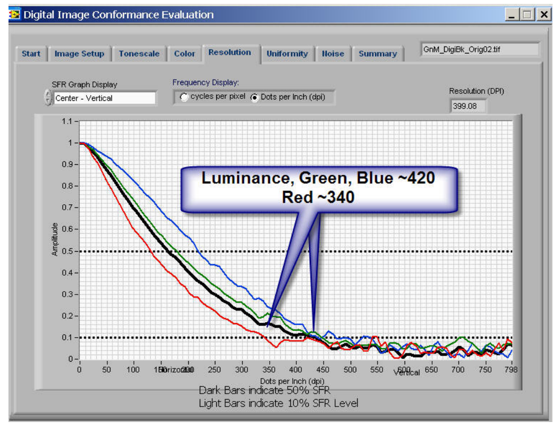


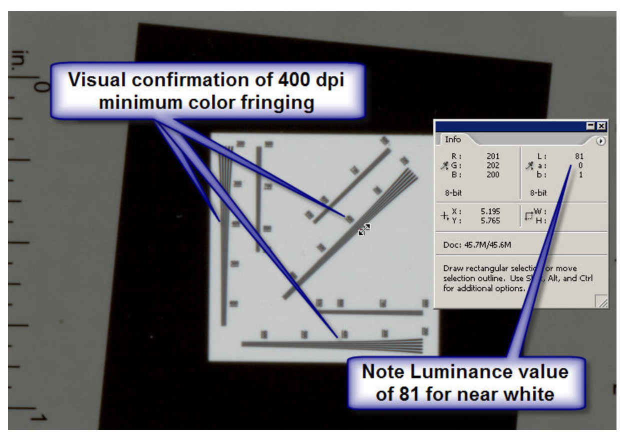
Fig. 6 – DICE Data Presentation, SOP Image

The RGB channel SFR curve tracking is reasonably good. All curves are smooth and parallel within reasonable distance from each other. The uniformity of the OECF curves is even better, so both *Fig. 4* and *Fig. 6* indicate there should be minimum color fringing.

Visual analysis – visual resolution

Fig. 7 provides visual confirmation of the software data presentation. The luminance value of the near white background is very low. In the original print, all three hyperbolic wedges can be read beyond the 400 dpi level. Color fringing does not appear within the wedges or around any of the target features until the image is displayed significantly larger than 1:1. (Almost all Library QA evaluations are done at 1:1.)

Fig. 7 – Visual Confirmation of DICE Data, Figs. 2 & 4



Software analysis – exposure variation

The SOP for this scanner included a “Digital Gain” setting of +16. “Digital Gain” is comparable to ISO sensitivity using proprietary algorithms and settings rather than ISO standard measures. The “Digital Gain” settings include 1, +4, +8, and +16. The scanner operators found that a digital gain of +16 was required to allow an aperture setting between f8 and f11 and the choice of the fastest scanning speed.

The use of digital gain to set scanner exposure presents a question of what would happen to image quality if other scanner controls were used to adjust exposure. Scanner operators can measure the effect on quality when other exposure control methods are used with the DICE target and software system. The scanner exposure was manually adjusted. The "digital gain" was set to 1, the scanner speed was set to double the exposure time, and the lens aperture was opened from f9 to f4. The resulting DICE OECF curve is presented in *Fig. 8* below. In this image, the white point has been moved up to a measured value of 242 – very close to the desired range of 243 to 250. The slope of the curve shows that contrast has been lowered. Patch three now shows a slight lightening and patch four is a visibly distinct shade of dark gray. Establishing a profile for these scanner characteristics might produce an image that is more suitable for document reproduction than for content research (although a derivative image file for research might be produced in post-processing. However, *Figure 9* shows the new image has lost significant resolution. The improved OECF curve (and improved color accuracy) came at a cost, probably as a result of adjusting exposure by opening the lens wide to f4.

Fig. 8 – DICE OECF Curve for a "Digital Gain" of 1.

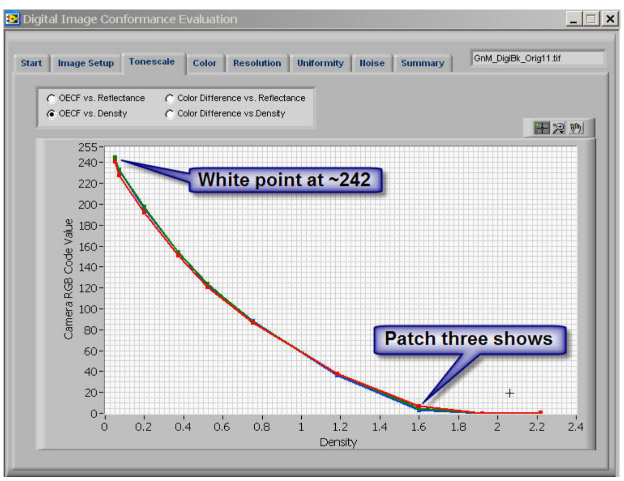
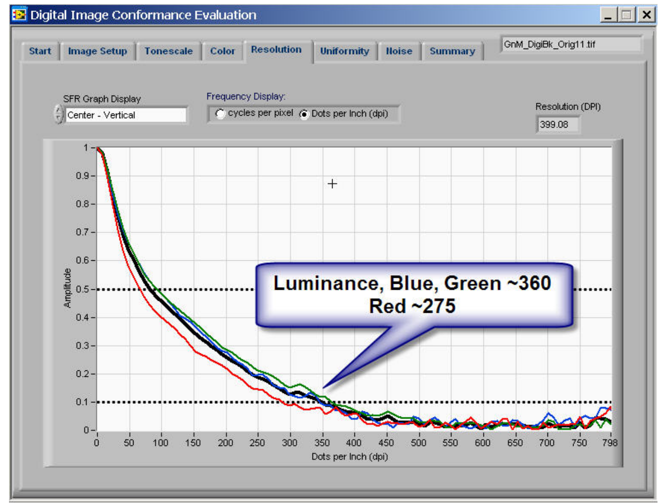


Figure 9 – DICE SFR(10) Curve for a "Digital Gain" of 1



Conclusion

The implementation of standards requires tools and quality assurance. The project shows how new targets and software provide some of the tools needed and how they may be used to create a baseline that is the foundation of a quality assurance program.

Author Biography

Barry Wheeler is currently a Digital Projects Coordinator in the Office of Strategic Initiatives at the Library of Congress. Previously he has worked as a web services specialist at the FDA Medical Library, and as a Professor of Library Science at The Catholic University of America. He holds a Masters in Library Science from Catholic University and a Doctorate in Instructional Computing from The University of Tennessee.