https://doi.org/10.2352/issn.2168-3204.2020.1.0.39

Refining the Theory-to-Practice Path for FADGI Still Imaging

Don Williams, Image Science Associates, Peter D. Burns, Burns Digital Imaging

Abstract

The Still Imaging guidelines report of the (Federal Agency Digital Guidelines Initiative (FADGI) continues to be a living document that has evolved from its initial digital imaging principles of over a decade ago. The guidelines have been adapted to the realities of day-to-day cultural heritage workflows. The 2016 version is a solid major improvement and has proven particularly useful in gauging digital imaging goodness. When complemented with training and evaluation tools, consistent compliant performance has been achieved. We highlight proposals for easier, less costly, and less frustrating ways to improve imaging performance and its evaluation under the FADGI guidelines – beyond benchmarking.

Introduction

The first version of the FADGI Still Image Guidelines [1] was completed in 2007 with the intent of helping to manage the range of digital image capture practices across cultural heritage institutes. In addition, the guidelines established nominal levels of image-capture performance. This was done using a 1-4 star rating scale. The current version was published in 2016.

The still imaging guidelines document continues to be a living instrument that has evolved from its digital imaging origins to accommodate the realities of day-to-day cultural heritage workflows. The 2016 version is a solid major improvement and has proven especially useful in gauging digital imaging goodness. With coaching, encouragement, and focused attention to detail many users, achieved reliable compliance, often with, e.g., high-speed sheet-fed document scanners.

In part, this is a testimony to an improved digital image literacy for the cultural heritage sector, articulated at the beginning of the last decade. [2] This objective and science-based literacy has certainly evolved and continues to do so. It is fair to say that no other imaging sector has such comprehensive objective imaging guidelines as those of FADGI, especially in the context of high-volume imaging workflows. While initial efforts focused on single instance device benchmarking, future work will concentrate on performance consistency over the long term. Image digitization for cultural heritage will take on a decidedly industrial tone.

As a community, we continue to learn and refine the practical application of FADGI guidelines in the preservation of meaningful information. Like rocks in a farm field revealed each year, new issues and errors with current practices call for refinement of methods and tools. Some are incidental, others need resolution. The goal of this paper is to highlight these and make proposals for easier, less costly, and less frustrating ways to improve imaging goodness through the FADGI guidelines.

Guidelines Proposals: Beyond Benchmarking

The motivation for this effort is to ease the use of and reduce the frustration with implementing the FADGI guidelines. This frustration is not so much on the measurement techniques themselves (all vetted ISO protocols), but rather the interpretation of the values that these techniques yield. As class instructors for FADGI-related training, we experience firsthand the questions and confusion about certain areas of the guidelines. [3] Some of the questions are easily answered while others expose areas needing clarification. We emphasize that the FADGI guidelines are malleable within reason and cost considerations and are not absolute.

We observe that several of the guidelines could be improved for ease of use and completeness. This is an opportunity. Here we discuss several possible changes to the imaging performance portion of the FADGI guidelines. Our suggestions are based on the need for technical rigor, and frequency of occurrence.

Fuzzy Limits

I frequently hear music in the very heart of noise
— George Gershwin, composer

No practical quantitative measurement technique is without some statistical variability. [4] Sometimes this variation arises naturally from the process we are observing, e.g. daily image exposure variability. In other cases, the variation is due to the way we sample some inherently varying parameter, e.g. pixel-to-pixel value over a nominally uniform area. In both cases, knowledge of measurement variability, or error, helps us interpret results. This is particularly important when compared with a specified test level results in an accept or reject decision.

So, there will be some margin of variation in the measurement system that needs to be considered when making judgments for different levels of FADGI compliance. This is often seen, for example, in reporting failure for a single target feature when it is out of specification by a small amount. We have seen how the current hard thresholds can be frustrating to users when a *fail* report is issued when the performance-difference threshold is exceeded by as little as 0.5%.

Rather than setting a 'hard' failure threshold, we propose a moderated fuzzy zone. This warning zone could alert the user that performance is marginal and needs attention. This idea is still fully within the spirit of the FADGI criteria. After all, the operative word is *guideline*. This concept has precedent with certain scanner manufacturers when testing their equipment [5].

Of course, multiple warning alerts do need to be managed. That is, how many warnings are too many warnings? We propose a simple three-violation rule. If for any metric category, three fuzzy warnings are activated then a failure for that test is recorded.

Otherwise, a conditional *pass* is awarded. We expect that margins of 0.5%-1.0 % are appropriate starting points for these limits.

Maximum percentile color error

Too often, very low *average* color errors are achieved for a particular target-capture, but the color metric portion of the color encoding test fails because the absolute maximum error guideline is exceeded by a few, less-important colors. It has been suggested that a percentile color error be used instead of a single-color error maximum. A percentile metric is already in use by BasICColor Input software and was the motivation for using this measure under the FADGI guidelines for assessing boundary condition errors in color encoding. We feel this is a good moderated approach and adopting a 90% maximum color error specification rather than an absolute maximum would be a much better choice. This would still allow for a ceiling level color error specification but reduce the influence of outlier color errors.

However, it may still be prudent to keep a maximum color metric to guard against wildly out-of-bound color errors. Perhaps unknown to most users is the intended utility of the ΔE_{2000} metrics. As an indicator of perceived color error, the ΔE_{2000} formula was validated to a maximum error of only about 6.0. So, when singular, high color errors in the teens and beyond occur this should be cause for concern when assessing color encoding accuracy. It is an indication of a possible underlying problem with; the color profile, image processing chain, or scanner/camera design.

Performance metric reporting context

Reporting a single goodness number for a complete performance category (e.g., tone scale) without providing more information on the extent of that goodness, or lack of it, is only marginally helpful. If, for instance, a *fail* for a single patch within a 12-patch tone scale category occurs, it would be helpful to know quickly, and immediately which patch needs attention. We propose a deeper fractional report to better help in remediation.

This can be done with simple graphics rather than a visual sorting of numerical table values. A candidate graphic demonstrating this is shown in Fig. 1 below. Colored bars proportional in size to the fraction of different star ratings passed help provide greater performance specificity. For instance, 47% of the white balance patches were satisfied for the 4-star level, 19% at the 3-star, and 33% for the 2-star. These items could then be expanded to reveal greater detail.

Transition to full colorimetric specifications

The use of the word *color* implies human vision. So, the best approach for any of the visible-range energy-related performance tolerances (e.g., tone scale, noise, white balance) should be human-vision related, i.e. colorimetric, rather than camera/scanner related, RGB. While the two can be related through color-space and ICC profile transformations, the fundamental approach should be colorimetric. This approach was not emphasized in the initial version of FADGI because most users at that time were only familiar with *RGB* data. *L*a*b** colorimetry was not part of the larger community's vocabulary or practice. With a decade of exposure to-colorimetric evaluation though, this change should begin. Recall, that ΔE_{2000} was always reported in the original *RGB* evaluations.

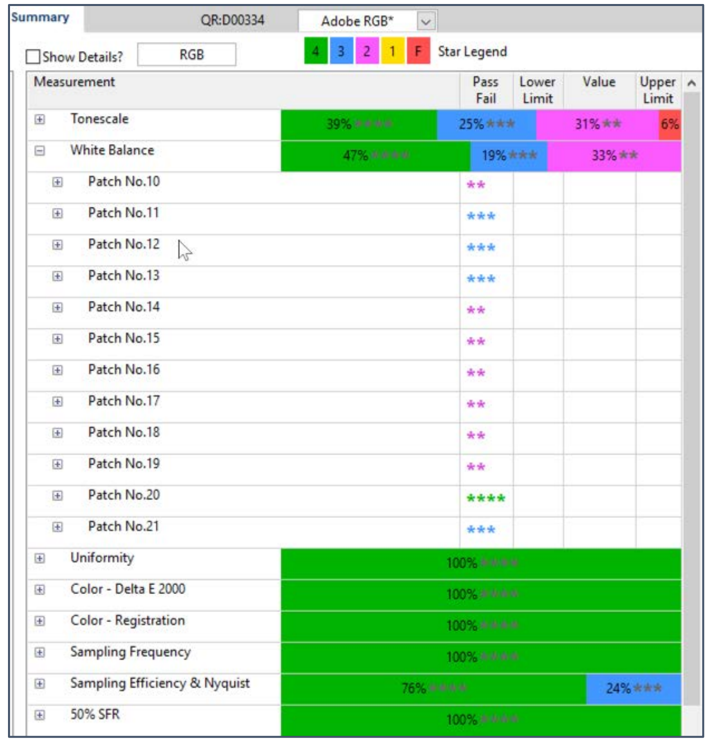

Fig. 1 – Color bar graphic for RGB performance reporting

A colorimetric approach will normalize the performance metrics to a human perception model, but the greater question remains on how this *RGB* to *L*a*b** transition is accomplished without greater confusion. We propose a dual *RGB-L*a*b** approach where both classes of metrics can be evaluated with a simple software toggle. Fig. 2 below shows the equivalent version of the same target scans but with a colorimetric evaluation report.

Film content specification

Current 4-star specifications for film content in the current FADGI document include 4000 pixels/inch (ppi) levels at 90% sampling efficiency. [6] This level is largely unachievable with normal scanning equipment and should be reconsidered with a more content-based approach. For instance, large format transparencies (glass plates, negatives, etc.) from the early 20th century are unlikely to have even 3000 dpi worth of information.

Many of the camera lenses and films of that era simply were not good enough to support that kind of spatial resolution. On the other hand, mid-20th-century professional cameras and films could likely achieve such information content detail. A more reasonable sampling frequency ceiling (i.e. 4-star) would be 3000 ppi, with corresponding rates of 2500, 2000, and 1000 for other star-rating levels.

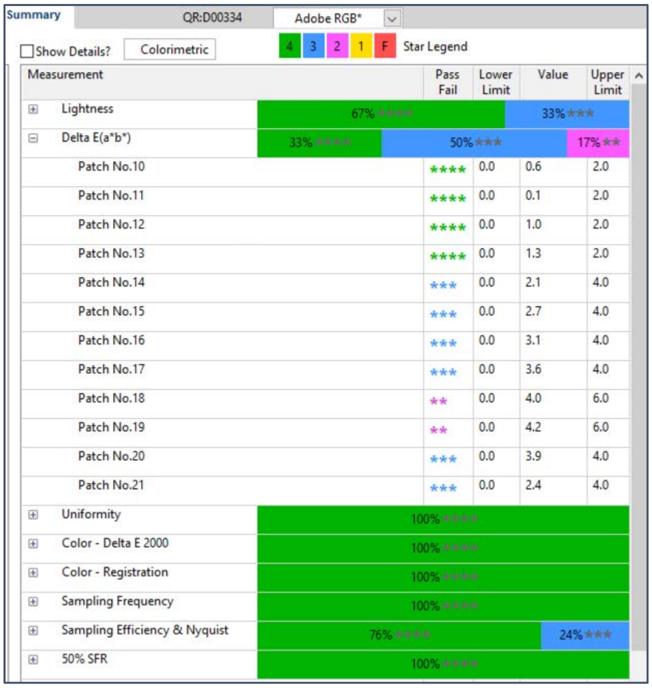

Fig. 2 – Color bar graphic for Colorimetric performance reporting

It is worthwhile recalling that the above values apply to smaller format negatives and transparencies (up to 4"x5"). While somewhat general, it is fair to say that formats larger than 4" x 5" require less resolution. This assumes that the larger format content derives from older negatives that pre-date the mid-20th century. While there are always exceptions, the camera lenses and films of that era were not of sufficient quality to support the kind of spatial resolution greater than 2000 dpi. This film size segregation is accurately reflected in the current guidelines.

Federal Records Category

FADGI guidelines were originally established for valuable cultural heritage content of public and private museums, libraries, and archives. We are now seeing a greater interest in adopting similar criteria for Federal or institutional records, especially temporary ones that are slated to be destroyed after being digitized. In the US, this is of greater focus with the US Federal Records Act at 44 U.S.C. 3302. Specifically, it states, 'the National Archives and Records Administration (NARA) will no longer accept paper/analog records for storage at a Federal records center after December 31, 2022' [7]. Similar requirements for other government departments and agencies (DoD, EPA, etc.) may be close behind.

The current FADGI guideline architecture can easily accommodate a *Records* category by defining relaxed performance criteria, where appropriate, for such content. For instance, the need for dropping tone-scale performance beyond optical densities of 1.6 may be reasonable for such a category since optical densities beyond this are highly unlikely for the designated content.

On the other hand, a greater focus may be required for establishing boundaries for scanning artifacts. Because of the volume of these records, most of them will be digitized on high speed sheet-fed document scanners. These devices are prone to deliver images with streaks, loss of texture, and other image artifacts. To date, though, we have seen examples of sheet fed document scanners and mobile devices [8] achieving 3-star level performance routinely. With the suggested improvements cited here a much larger community of practice is possible.

Boiler Plate Restraint*

Trouble is easily overcome before it starts

— Lao Tzu, philosopher

We recognize the cost implications of over-specifying job digitization job requirements. Doing so is becoming common as the guidelines are being adopted more frequently. It is indeed tempting to specify a 4-star level requirement because the implication is high consistency and quality. And, of course, everybody's collections are special. But the cost levels for such are very high and it takes on a 'Why use lead when gold will do' attitude.

Such 4-star specification templates in digitization contracts can be unnecessary but for exceptional content. Perhaps the greater risk in asking for such quality is the user's ability to audit such performance. In the end, one may be paying for 4-star performance but unknowingly receiving something less. Remember, while a singular 4-star performance event may be demonstrated the greater task is maintaining that performance over an entire project's history in a high-volume workflow. We therefore encourage issuers of digitization contract specifications to restrain from requiring a blanket 4-star requirement but instead start with 2- or 3-star requirements. There are many digitization projects at those levels currently that are doing very well. As noted, the greater need may be to ensure such performance levels are maintained over the course of a project, *consistency is the key*.

Achieving 4 star, but how?

Great things are done by a series of small things brought together

— Vincent van Gogh, painter

Yes, 4-star level can be achieved but it requires attention to detail. Over the years we have noted several factors that can help users achieve these levels. We list several of these below.

Custom Color profiling: Creating custom color profiles can be very helpful since it avoids assumptions on lighting, target colorants, and image processing used to create batch color profiles. Having good white balance on the input target image before color profiling is beneficial. [10]

* "Boilerplate documents are commonly used for efficiency and to increase standardization in the structure and language of legal documents …" [9]

Target measurement: Rather than using batch color reference data for a test target, we use custom-measured values for each individual target. All current color profile engines now allow for the ingest of such custom data.

Moderated post-processing: There is nothing wrong with a moderated amount of post-processing to align image data with the guideline specifications. Some practitioners frown upon this as if the data first out of a camera or scanner is virgin in some way. It is not. A healthy amount of image processing has already been done before the first image is delivered to the user. This applies to RAW data too but to a lesser extent. Such processing can be applied via an Application Programming Interface (API) or through a Command Line Interface (CLI) for more transparent workflows.

Frequent audits and operator training: Both items are big factors in reducing variability in digitizing workflows. If a single device calibration will hold over an extended period or that untrained operators will catch mistakes is naïve. 'What could go wrong?' is not a good model to adopt with such high-volume workflows for cultural heritage imaging.

Conclusions

With more than a decade of experience of FADGI testing, user feedback, and training we are suggesting refining FADGI testing tools and procedures. It is becoming clear the tools and methods for compliance testing do work, but improvements and user training are called for. The content-specific architecture of the guidelines makes them flexible and easier to adjust for differing applications.

In the future, we foresee these guidelines extending to multi-spectral imaging and other advanced imaging techniques specific to cultural heritage imaging.

References

[1] FADGI Still Image Working Group, ed. T Rieger, US Library of Congress, http://www.digitizationguidelines.gov/

[2] D. William and P. D. Burns, A Decade of Experience with Digital Imaging Performance Guidelines: The Good, the Bad, and the Missing, Proc. IS&T Archiving Conf., 165-169, 2017.

[3] P. Conway and D. Williams, Enhanced Education for Better Imaging Practices, Univ. Michigan website, (2011). http://www.academia.edu/2796022/Enhanced_Education_for_Better_Imaging_Practices_A_Case_Study_at_the_University_of_Michigan

[4] P. D. Burns, Estimation Error in Image Quality Measurements, Proc. SPIE vol. 7867, 2011.

[5] Personal Communication, Volker Jansen, Technical Director and Manager R & D.

[6] P. D. Burns and D. Williams, Sampling Efficiency in Digital Camera Performance Standards, Proc SPIE 6808, 680805, 2008

[7] https://www.federalregister.gov/documents/2019/04/10/2019-06916/electronic-records-management.

[8] P. D. Burns and D. Williams, Going Mobile: Evaluating Smartphone Capture for Collections, Proc. IS&T Archiving Conf., 2016.

[9] Investopedia website ,https://www.investopedia.com/terms/b/boilerplate.asp

[10] D. Williams and P. D. Burns, Targeting for Important Color Content: Near Neutrals and Pastels, Proc. IS&T Archiving Conf, pg. 190-194, 2012.