Color Fidelity Test Methods

Michael Stokes
Hewlett-Packard Company, Boise, Idaho/USA

Tom White
Microsoft Corporation, Redmond, Washington/USA

Abstract

While there has much published about color management and methods to achieve accurate color, little practical information has been published about test methods to verify or validate that color fidelity aims have indeed been achieved. This paper discusses why testing methods are important, describes a basic list of considerations for establishing test methods and finally details four practical color fidelity test methods.

Importance of Testing Methods

One of the ongoing customer complaints against color management solutions over the last ten years is that they often simply don't produce good visual matches in real world situations. Given the number of recent books[1,2,3,4,5,6,7,8,9,10,11,12,13,14,15], articles[16,17], web sites[18.19,20.21.22.23], tutorials,* consortia[24,25,26,27,28,29,30]and conferences[31,32,33,34,35] that address the communication and management of color information, this might seem somewhat surprising. The simple answer is that cross-technology and cross-viewing conditions color reproduction in an open network is a very complex problem. At its most basic level, this problem is an attempt to understand, model and simulate a large portion of the human visual cortex. Despite this complexity, some isolated installations and carefully controlled demonstrations have indeed achieved good success, but seldom, if ever, do these reflect real world situations for mass markets in open networks.

These ongoing complaints are due to a variety of real world problems and situations. Some of these complaints are due to different software applications that display and print unacceptably different colors when representing Pantone colors. It is still difficult to get scanners, displays, computers and printers from the same manufacturer (much less different manufacturers) to produce acceptable color matches without additional and sophisticated intervention. It is also frequently difficult to get software applications from the same vendor to produce acceptable color matches while interacting with other applications from the same vendor. Finally, it remains unclear where and how color management is performed and what device or software directs this performance in such a way to insure that multiple redundant or conflicting color transformations are not executed. This is by no means a complete list of the problems involved currently in color management "solutions," but it does provide a representative sampling of some of the most problematic objections. Until these fundamental objections are overcome, it will continue to be difficult to achieve adoption of color management solutions in mass markets.

The creation of a viable color-imaging product is a balance of many often conflicting product requirements, including cost, performance, ease of use, size, flexibility, reliability and color image quality. While color image quality is one of the most complex requirements, it remains only one of many user requirements. It often consists of many underlying aspects such as resolution, color depth, spatial uniformity, color accuracy, color precision and dynamic range to name a few. Given this complexity, it is easier to focus on other product features as both a vendor and a reviewer. Despite this, some reviewers have tried valiantly by showing a series of different print sample representations in various magazines., These reviews are obviously limited by some of the very reproduction issues that are being reviewed and only show at most a limited number of the aspects that make up color image quality. It is the authors' belief that unless and until a set of simple color fidelity test methods can be ubiquitously established, little progress will be made in overcoming the very real and difficult problems that currently limit color management solutions.

One simple example of the potential impact of problems caused by these limitations is illustrated when one examines electronic commerce. Traditionally a product vendor creates a printed catalog of their inventory and distributes this catalog to a large number of potential customers. The printing and distribution process are very expensive and the actual customers represent a relatively small portion of the entire distribution list. Finally, the number one customer complaint of catalog shopping is poor color matching between the catalog and the final product. Web-based catalog shopping is much more cost-effective for the product vendor in that it eliminates both the printing and the distribution of the catalog. Yet to be successful in replacing traditional methods, it appears that it must be both more convenient and reliable than traditional methods. This means that the color complaints must be dramatically reduced. Unfortunately, whereas a traditional catalog is created in a relatively closed color reproduction workflow,

---

* Like most conferences, this IS&T Conference has annually sponsored a number of tutorials. Additionally, in depth tutorials are available from a variety of individual consultants.

web-based catalogs exist in a very open and often ill-defined computer network. Properly managing the unambiguous communication of color information in an open network, across application, utility, device and operating system vendors seems to be a fundamental requirement in resolving this problem. While there are clearly other factors involved in establishing successful electronic commerce solutions, a weak or limited color management solution will pose a significant hindrance to the process. This is just one example of many that could have been posed.

One of the roadblocks to overcoming these problems is that there is not a clear or simple method to show quantitatively that a color reproduction system isn't working acceptably. As with any technical problem, it is much easier to fix it or improve it if one has some measure of whether and how much it is broken and in what way it is broken. The entire industry of statistical quality control rests upon this basic premise and has been well documented for decades. Currently there does not exist any widely accepted method or set of methods to objectively and quantitatively evaluate color management solutions. This paper provides the basic description and philosophy for such a set of methods. A full set of implementation details based on this paper will be published in the near future.

Considerations in Establishing Test Methods

To even begin to establish test methods to address the problem described above, it is necessary to understand, consider and prioritize the many factors that affect color fidelity. It is common for a particular vendor or product reviewer to try and consider each device or application in isolation. Although this approach might seem reasonable, it often masks the more difficult problems of integrating the device or application into a real world workflow. Another common approach that can provide very misleading information is summarizing a device's color fidelity with an average CIELAB delta E* of some particular pictorial image. While this approach does provide some relevant information, it rarely provides a robust view of the color fidelity issues of an application or device. In addition, it compounds known weaknesses in CIELAB, such as a perceptual blue-purple non-linearity, directly into the testing method. This could cause devices that actually provide perceptually superior color fidelity to be at a disadvantage to devices that are optimized to CIELAB. While no one single number can ever adequately provide a robust evaluation of color fidelity, some reporting processes must be extremely concise when evaluating many different aspects of a product. In circumstances where a single summary number is provided, it is recommended that a clear explanation of how this number was derived be available.

One of the possibly obvious goals of color fidelity test methods is to reflect and test the multiple software and hardware paths that underlie the most common use cases end users implement. Color reproduction systems today are combinations of hardware, software, utilities and operating systems. Each of these components provides a very wide variety of options that directly affect color reproduction. A very small subset of these options includes different paper, resolution, interpolation, color space, device drivers and profiles. While the actual number of options varies with each vendor product, it is clear that reporting, much less testing every combination of all of these options is simply impossible.

It is impractical to test all possible combinations of devices and software. Therefore it is recommended that the manufacturer's recommended default setup be used in most cases. An exception would be in the case of evaluating a particular workflow such as CMYK SWOP prepress workflow. This also applies when choosing the non-targeted aspects of the reproduction workflow. For example, when evaluating a printer, choose the most common components when constructing the rest of the system. By choosing the most common use cases, applications or devices as reference standards, it becomes practical to establish a stable environment in which to test individual components. Once a test environment is established, try and keep it as stable as possible in order to provide some level of objective comparison over time. Such an environment allows one to test individual applications, CMMs, profiles, devices, and calibration utilities in an objective manner.

The one exception to this approach of establishing a stable default environment is when one is testing robustness against a wide variety of systems. For example, when developing new printer products, it is necessary to test against a wide range of software applications. By maintain as stable of an environment as possible, it becomes easier to isolate compatibility problems in the development process.

An additional aspect of color fidelity test methods is to appropriately target these methods to their users. This paper describes four methods that trade off ease of use with informational accuracy and complexity. It is both futile and unreasonable to arbitrarily mandate that end users implement a full-blown psychophysical testing process. It is just as problematic to assume that a simple visual assessment is adequate for device manufacturers in the product development process. While a simple visual assessment will not provide the robust information that a complex colorimetric assessment will, it may provide all information required by a user. The series of methods described below progressively increase in complexity and resource requirements. This increased complexity and resource requirements is directly associated with an increased accuracy and depth of information. For example, the first method is a simple visual assessment that provides only pass/fail information against a known tolerance. The complex colorimetric assessment not only provides a summary number, but breakdown of quantitative color differences by hue, chroma and lightness in varies color quadrants. Therefore it is critical that the test methods appropriately target the needs and resources of their users.

Method 1: Simple Visual Assessment

The first method is a simple test document that exercises a number of software paths. Such a document is described along with simple guidelines for acceptability. These guidelines include straightforward instructions such as "make sure red is not orange, blue is not purple, and there are no cyan dots in yellow." This method is ideal for quick visual tests that are a small aspect of larger test suites. Such suites are common in application and device development schedules.

The document should be a combination of text, colored text, charts, clip art and raster images to ensure robust testing of processing paths through the system. It is common for applications, utilities, operating systems and devices to treat different object types differently. If care is not taken during the development process, it is easy for these different paths to produce unintentional differences. For example, a simple Microsoft Word or Adobe PageMaker document could be created to meet these requirements. Whatever application or document format is chosen, it is critical that the color descriptions of the objects in the document be well defined.

There have been attempts to provide reasonably stable standard images and documents to base these assessments upon. ISO TC130 WG2 and the Committee for Standard High Precision Pictures (SHIPP) [36] have both developed a set of standard pictorial images and measurement patches to enable color fidelity assessment of pictorial images. The ISO Standard Color Image Data (SCID) targets the graphic arts market and provides CMYK data. The SHIPP images are based on ITU-R BT.709-2 and sRGB data and contain device-independent color data. Finally, more complex documents can be created to test a wide range of issues including color fidelity, performance, resolution and cost-per-page. Spencer and Associates provides a good example of this type of work[37].

Both SHIPP and SCID data sets attempt to unambiguously describe color information and how to use it. This unambiguous explanation is critical when attempting to integrate standard color objects such as images into a test method. There are three basic aspects of color that are required to unambiguously describe color in a manner that is useful for test. These are the reference device, viewing condition and observer. With each of these descriptions it is necessary to also provide the underlying models. Device-independent color spaces are based on the human visual system, so the device and observer are identical. This does not eliminate the need to provide a clear description of the viewing conditions and the underlying appearance model to transform into and out of these viewing conditions. For example, when using CIELAB or CIEXYZ, it is critical to provide descriptions of both what the viewing conditions are (surround, white point, luminance level, etc.) and how these conditions are modeled. Many standards provide only the parameter levels and not the underlying models. This approach is misleading at best. It is comparable to speed limit signs that say "55" but don't indicate if it is in miles-per-hour or kilometers-per-hour. Providing an ambient white point without describing the chromatic adaptation or color appearance model is a good example of this problem in the color field. It assumes that any chromatic adaptation or color appearance model is viable. This is simply untrue and many studies have proven that such models are significantly perceptibly different from each other[†], especially in saturated graphics such as corporate logos. Another common example of this problem is when a 'gamma' is specified such as 2.2. Without providing an underlying model that might or might not include an offset, gain and flare component, it is unclear what the actual implementation of this parameter is. For example, ITU-R BT.709-2 has an exponential value of 2.4 and this parameter is often cited as the 'gamma' for HDTV, ignoring the significant offset component. When the offset is factored in, a simple power function (with zero offset and a gain of one) fits the exact same recommendation with a 'gamma' of 2.2. Which is correct? Both are, within their own context. One objection raised to clearly specifying a color appearance model is that it is either proprietary or still under active research. This is a superfluous and misleading argument since the parameters provided are still based on some real world implementation.

In conclusion, a simple visual assessment document provides very high ease of use with a rough assessment of acceptability.

Method 2: Complex Visual Assessment

The second method is a more complex, but is still a reasonable visual test that uses a page of solid reference colors (some in gamut, some out of gamut) and the code to produce this on a printer. It can also contain the information in the simple visual assessment document. This test provides direct, visual comparison with known tolerances also included in the test page. It is also easily extended to provide tests for scanners and display. Such a test is ideally suited for product reviewers or test engineers or reviewers who do not have easy access to measurement equipment or have significant time constraints.

The actual assessment consists of a series of reference colors and tolerances that make up a visual template. Like the "variations filter" in Adobe Photoshop, the target color is surrounded by patches with varying hues and chromaticities at a specified tolerance. In addition, patches of greater and lesser lightness are also provided. By using a template with the target patch position cut out, the user can simply overlay the template onto the target color print and visually assess whether the target patch color is within the visual tolerances provided.

Since simple patches are used, it is critical to adequately sample the device's color gamut and to separate in-gamut colors from out-of-gamut colors. There is an implicit weighting of gamut information by the choice of target colors.

A similar method can be applied to scanners with the use of a digital template and a calibrated monitor. An alternative scanner method is to provide a digital template and an application utility that compares the scanned results against the target results.

The advantage of this method over the simple visual assessment method is that an unambiguous acceptability

---

[†] The numerous articles in previous IS&T/SID Color Imaging Conferences on color appearance models and white point adaptation models clearly illustrate these differences.

tolerance is inherent in the evaluation and additional qualitative information on which part(s) of the color space are acceptable and which are not and in what direction.

Method 3: Complex Colorimetric Assessment

A third method uses a test target similar to that in the second method. It is of course possible to provide additional patch targets or use standard patch targets such as the ISO TC130 WG2 targets or uniformly sampling the device color space or a device independent color space. Color differences are computed between the original target and the output. These are summarized using Pointer and Hunt's color reproduction index methodology[38,39,40]. This produces not only a summary average delta E*, but also provides more detailed information, such as differences in lightness, chroma and hue and which direction the printed output differs from the reference target in each of these dimensions. By including both in-gamut and out of gamut colors, the gamut mapping issues can be summarized separately. Both this test method and the previous can be modified to test different processing paths.

Additional information can be derived and reported using this method. One example is metrics concerning gamut volume and shape. There are many issues involved in comparing color gamuts of different devices, especially between capture, display and print technologies. These include complex issues such as either having similar viewing conditions or a standard method for compensating for viewing conditions. They also include issues that appear simple, but are in reality quite subtle and complex, such as what is white and what is black in the gamut comparison. Finally it is necessary to not only report a single number, such as gamut volume in a uniform color space and a typical viewing condition, but metrics that describe the shape of the gamut in a reasonable manner. This might include gamut circumference as various levels of grey or volume mismatch to standards gamuts such as SWOP or sRGB. Gamut metrics are only one example of many other descriptive metrics that can be derived from this method.Such a test is ideally suited for product color engineers who want to isolate and improve any color defects or artifacts in a product. While this method requires the use of a color measurement device, it provides much more detailed quantitative results about the color fidelity of the product being evaluated. In addition, there are many automated measurement stages that simplify the time consuming task of measuring individual patches.

Method 4: Psychophysical Assessment

The final method of assessment is a full-blown psychophysical evaluation. A series of test documents are reproduced. A standard viewing condition environment is created that is representative of the target user market. A pool of color normal observers is then recruited. Using standard psychophysical methods, a variety of evaluation questions can be submitted to the observers. Statistical analysis of the results provides superior information that is independent of any limitations of color appearance, color space or color difference models.

This method eliminates any dependence upon potential flaws in color appearance models in the previous method. It also provides detailed information on what aspects of the product are strong and weak with respect to color fidelity. It is not uncommon that the most valuable information is found in the general comments by the observers that are not part of core evaluation questions. This method also provides the capability to determine the priority of color fidelity issues in comparison with other product requirements. Unfortunately setting up, administering, and analyzing psychophysical experiments is a very intensive and time consuming process has often has a significant amount of statistical variation. While ideally this reflects the true end user, it can be difficult to extract clean conclusions.

This brief summary of psychophysical testing cannot do justice to this complex field. There are extensive references available for those interested in this method[41,10,12,42].

Conclusions

In conclusion, this paper describes actual color fidelity problems in manufacturing and development environments and practical solutions that were implemented to successfully address those problems. This approach and context can be extended or modified to address other related color fidelity test issues.

References

1. Michael Kieran, Desktop Publishing in Color, Random House, Inc., New York, 1991
2. Dan Margulis, Makeready: A Prepress Resource, MIS:Press, New York, 1996
3. David Bann and John Gargan, Colour Proof Correction:Question and Answer Book, Phaidon Press Limited, Oxford, 1990
4. Tony Johnson, Colour Management in Graphic Arts and Publishing, Pira International, Surrey, 1996
5. Charles A. Poynton, A Technical Introduction to Digital Video, John Wiley & Sons, 1996
6. Harry Mathias and Ricard Patterson, Electronic Cinematography:Achieving Photographic Control over the Video Image, Wadsworth Publishing, Belmont, 1985
7. Apple Computer, Inc., Advanced Color Imaging on the Mac OS, Addison-Wesley Publishing Company, Reading, 1995
8. Michael Kieran, The Color Scanning Success Handbook, DPA Communications, Toronto, 1997
9. Roy Hall, Illumination and Color in Computer Generated Imagery, Springer-Verlag, Berlin, 1989
10. Mark D. Fairchild, Color Appearance Models, Addison-Wesley, Reading, 1998
11. W. N. Sproson, Colour Science in Television Display Systems, Adam Hilger Ltd., Bristol, 1983
12. Richard S. Hunter and Richard W. Harold, The Measurement of Appearance, 2nd Edition, John Wiley & Sons, New York, 1987

13. R. W. G. Hunt, The Reproduction of Colour in Photography, Printing & Television, Fountain Press, Hertfordshire, 1987
14. Edward J. Giorgianni and Thomas E. Madden, Digital Color Management, Addison-Wesley, Reading, 1998
15. Rolf G. Kuehni, Color:An Introduction to Practice and Principles, John Wiley & Sons, 1997
16. Color Engineering (private communication with Max Salzman)
17. Color Research and Application, John Wiley & Sons, New York
18. Hewlett-Packard Company, http://www.sRGB.com
19. International Electrotechnical Commission, TC100, PT6-1966, http://w3.hike.te.chiba-u.ac.jp/IEC/100/PT61966/
20. Committee for Graphic Arts Technologies Standards, http://www.npes.org/standards/cgats.htm
21. Charles Poynton, http://www.inforamp.net/~poynton/notes/links/color-links.html
22. Color & Vision Research Laboratories, http://www.cvrl.ucsd.edu/index.htm
23. Ian Kelly, http://mcs.open.ac.uk/ik7/colour.html
24. International Color Consortium, http://www.color.org/
25. Graphic Arts Technical Foundation, http://www.gatf.lm.com/
26. Japan Electronic Industry Development Association, http://www.jeida.or.jp/guide/gaiyou/index-e.html
27. Video Electronic Standards Association, http://www.vesa.org/
28. World Wide Web Consortium, http://www.w3.org
29. Virtual Reality Meta Language Consortium, http://www.vrml.org/
30. Digital Imaging Group, http://www.digitalimaging.org/
31. IS&T Conferences, http://www.imaging.org/conferences/conferenceindex.html
32. International Colour Association
33. ACM SIGGRAPH, http://www.siggraph.org/conferences/conferences.html
34. Inter-Society Color Council, http://www.iscc.org/
35. International Commission on Illumination, http://www.cie.co.at/cie/home.html
36. Koichi Sakamoto and Hitoshi Urabe, Standard High Precision Pictures: SHIPP, Proceedings of the Fifth Color Imaging Conference, IS&T/SID, Scottsdale, 1997, p. 240
37. SpencerLab Digital Color Laboratory, http://www.spencer.com/TestReports.html#Software
38. Michael Pointer, Measuring Colour Reproduction, *J. Photogr. Sci,* **34,** p.81 9186)
39. C. A. Wood, M.R. Pointer and G.G. Attridge, The Application of Colour Reproduction Indices to Photographic Reflection Prints, *J. Photogr. Sci,* **35,** p. 66 (1987)
40. G.G. Attridge, C. Leverton and M.R. Pointer, Measured Colour Differences and the Acceptability of Photographic Colour Prints, *J. Photogr. Sci.,* **43,** p.169 (1995)
41. Franc Grum and C. James Bartleson, Optical Radiation Measurements: Volume 4: Visual Measurements, Academic Press (1980)
42. Deane B. Judd and Gunter Wyszecki, Color in Business, Science and Industry, 3rd ed., John Wiley & Sons, New York (1975)