

Test Image Design Guidelines for Color Quality Evaluation

Gary G. Field
California Polytechnic State University
San Luis Obispo, California, U.S.A.

Abstract

The design or selection of test images are important aspects of the color quality evaluation process. Images may be used to assess process performance, evaluate process consistency and to provide process feedback information. They may have analytical or pictorial structures. Images for process optimization or subjective quality assessment, however, have certain drawbacks when used for color reproduction quality research. Color Mondrian or "quasi-Mondrian" images used in conjunction with related pictorial subjects are suggested options for color quality research purposes. General guidelines for test image selection, design and evaluation are presented.

Evaluation Objectives

There are three broad color quality evaluation processes that are of general interest to the imaging industry. The first is the performance assessment of equipment and materials that are used in the color reproduction process. Scanners, printers, output media and monitors are routinely evaluated to determine which is most suitable for a particular application.

The second class of color quality evaluation process concerns the optimization of a given set of color reproduction equipment and materials. This latter objective is a production evaluation process whereas the first objective applies to the equipment/material purchase or manufacturing process.

The third color quality evaluation process is that used by the research community to test color appearance models and explore aspects of optimal color reproduction theory. The objectives of such research include standards development, improved procedures and equipment and materials refinement.

In order to satisfy the objectives of each class of color quality objective, special test images are usually employed. Such images must be carefully chosen and/or designed in order that the evaluation process is efficient and valid.

Types of Test Images

Test images may be discussed under three broad headings: process performance, process consistency, and process information.[1] Some images may be classified in more than one category, but most images have a specific task that is exclusive to one category.

The performance assessment type of test image refers to the absolute or ultimate quality that may be achieved by a particular technology or material. Such assessments as resolution, color gamut, tone range, random or periodic patterns, and such physical factors as dimensional accuracy are also considered. Assessments may be made by physical measurement, from analytical types of test image, or by psychophysical measurement from pictorial images.

Consistency evaluation images are usually employed in production to monitor the ongoing image manufacturing process (e.g., photofinishing, copying or printing). These images are always non pictorial and are usually composed of small target fields arrayed in a format that is suited to rapid evaluation.

Process information images are feedback signals used to help program image processing systems with the characteristics of the recording media and output device combination, or with the characteristics of the input media and input capture device combination. The test images used for color management purposes are of this type and, in general, have an exclusively analytical structure.

Analytical Images for Process Optimization

Test images may have either an analytical or a pictorial structure. Each have specific strengths and weaknesses that dictate their preferred uses.

The kinds of image quality factors that are evaluated via test images with an analytical structure include: mottle and other patterns (e.g., moire), evenness, optical distortion, register, color balance (gray field analysis), Dmax, color gamut, additivity failure, proportionality failure, Dmin, dot gain, resolving power, and colorant strength.

Densitometers, spectrocolorimeters and other quantitatively-oriented systems of measurement are used to monitor or evaluate the analytical type of image. Image elements are arrayed in strip format for consistency evaluation, in rectangular format for feedback evaluation, and in full-page format for the performance evaluation of such qualities as evenness, mottle, and optical distortion.

Analytical images are usually purpose-designed for a particular class of equipment type and format. Some images (e.g., IT8.7 and SWOP) have a quasi-standard format. In either case, measurement technology, size restrictions,

chosen quality analysis factors, and equipment or material characteristics must be considered by the test image designer during the design and layout process.

Pictorial Images for Quality Assessment

Such factors as tone reproduction, saturation compression, sharpness, graininess, and overall picture quality are best judged via a series of pictorial images. In general, such images are used to assess the system optimization setup operation, but they may sometimes be used to good effect for assessing the performance characteristics of a particular system.

Test pictorial images should include those with low, medium, and high key tonal distributions, those with saturation levels ranging from low to high for all important hues, those with subtle tone and color transitions, and those with fine detail and sharply defined objects. Large, smooth areas should be considered for graininess assessment purposes and neutral grays should be used for color balance assessment purposes. High density range images with contrasty lighting and low density range images with diffuse lighting will also provide useful insights into imaging system performance. Test images should be available in a range of formats (e.g. 35mm to 8” x 10”) and in both transparent and reflective modes.

The subject matter in pictorial test images must be chosen with care. In general, images should be natural in terms of appearance (no contrived setups) and neutral in terms of content (no scenes or objects that are likely to bias observers for or against a particular image).

The broad classes of test image subjects may be identified as consumer images or professional images. Subjects for consumer images include portraits (several different ethnic groups), landscapes (many places and lighting conditions), pets, possessions (cars, houses), and seasonal celebrations. Professional images, which reach their zenith in the advertising field, include portraits, automobiles, food products, clothing, industrial and office equipment, travel-inspired landscapes and scenes, real estate, and sporting goods. Medical, forensic, macro/micro, aerial, astronomical, and scientific fields use specialized subjects that, ideally, should be reflected in the test images for those respective applications.

Pictorial images are assessed by panels of observers via psychometric scaling techniques. The statistical validity of such tests may be compromised by the content of the image;[2] therefore, test images must be chosen with great care.

In cases when pictorial images are being used for assessing a particular system performance quality (e.g. tone reproduction) a simple image is usually preferred in order to exclude the cross influences of other color or image quality factors. All-inclusive quality metrics are not usually reliable enough for evaluating overall performance, especially when specific image qualities may be of particular interest.

All-inclusive image quality metrics may be used, in some cases (e.g., consumer photofinishing or newspaper editorial color) for assessing system optimization setup operations, but in other cases (e.g., graphic arts advertising image color approval or professional photofinishing) the optimum will be established by a single judge (the customer) after a sometimes-lengthy iterative process. In the latter case, pictorial test images have little value because the intent is not to establish the best average setup conditions for a range of images; rather, it is to establish the perfect setup for a given image.

The major drawback with pictorial images is that it is usually impossible to make consistent quantitative measurements from the image. Variations in image detail will produce variations in measurement, and the inclusion of a test target within the scene destroys the naturalness of the image. This limitation on the use of pictorial images is of no concern when optimizing the setup of a given color reproduction system because such judgments are usually subjective. The color reproduction research field, however, prefers the objectivity of quantification.

Research Images

Color reproduction research work presents special problems for test image designers. Pictorial images introduce distortions based upon content whereas purely artificial images (e.g., the Macbeth ColorChecker[3]) are not representative of real images.

The Macbeth ColorChecker deserves special comment because of its extensive use in color reproduction studies. It has the dual advantages over pictorial images of being content free and of containing areas that may be readily measured by most instruments. The inclusion of gray scale, gamut colors, and a selection of "interest area" colors (e.g., skin tones) makes this target quite suitable for comparing the performance of various imaging processes. Colorimetric comparisons of trial images, however, are of limited value when the goal of the reproduction process is something other than facsimile reproduction.

Practical color reproduction objectives are usually based upon an "optimum" strategy that involves making preferred color changes, correcting unwanted distortions in the original, and making compromise tone and saturation compressions[4]. The exact nature of optimum reproduction strategy depends upon the structure and importance of the elements in the image: high key vs. low key, or merchandise color vs. decorative color, for example.

A promising form of test image for color reproduction research appears to be the color Mondrians that McCann[5] made extensive use of during his color vision researches. Mondrians are content free and consist of uniform tone areas that may be readily measured and evaluated. A selection of Mondrian images may be used as surrogates for a wide range of image types. The subsequent reproductions may be ranked by Pointer's Colour Reproduction Index (CRI) prior to using compatible pictorial images in confirming visual assessment experiments.

Image Design and Selection Practices

A great number of standard test images (both analytical and pictorial) are available for purchase from research associations and other organizations. In general, standard images are preferred to custom-designed images because they are less expensive, are made to demanding standards, and make it easier for test image comparisons to be made between different companies and organizations. Standard images do, however, have some drawbacks. Some appear artificial, because of the attempts to create universal multi-purpose images, and others do not necessarily incorporate the specific colors or details of particular interest to the evaluator.

Custom design of analytical types of test image is not uncommon because of both their relative simplicity, and the many different control requirements for different types of imaging systems.

Pictorial images are more difficult to produce; therefore, the use of stock images may prove to be a useful alternative. Custom-designed pictorial images usually require the services of a professional photographer. Such images can become dated (hair style and clothing fashion changes) but for specialized research or testing purposes, more current reshoots can be justified. In order to evaluate the full range of color and image quality factors, anywhere from 5-10 different pictorial images could be required.

Mondrian images have to be custom designed by researchers to represent a range of tonal, hue, and saturation distributions within common images. Alternatively, low resolution scans and digital image processing techniques could be used to construct "block pixel" abstractions from actual images. These "quasi-Mondrian" designs which are really stylized versions of the image in color block format, may prove to be more useful than the classic semi-random Mondrian. Further research is needed.

Evaluating Test Images

Analytical test images may be evaluated by visual and instrumental means. An efficient design, from the visual viewpoint, displays the image elements in an intuitively logical progression or layout. Ultra sensitive elements (including those with accept-reject features) are included to provide early warning of process variability, and to characterize the performance limits of the imaging system.

Efficient analytical test image design, from the instrumental evaluation viewpoint, is realized when the required information is extracted from the image with the minimum effort. The data processing stage must be considered along with the measurement process when making the evaluation. In the case of look up tables (LUTs), for example, the test image design will be influenced by the interpolation program that is used within the LUT.

The evaluation process for pictorial image design is somewhat subjective. The actual distribution of tone and color values within images may be measured, as may such other factors as density, resolution, and granularity. The selection of ideal target values for a given application may have to be based upon the empirical evaluation of multiple trial images. The suitability of the subject matter within a pictorial target image may be assessed via observer panels. Statistical tests are used to reject images that test significantly negative or positive to any sub group of the observer population, or to the population as a whole.

The success of Mondrian or quasi-Mondrian designs is related to how well their assessment predicts psychophysical evaluations of real images. Mondrian elements must be sufficiently large to enable measurements to be made.

Conclusion

Test images are crucial elements in the processes of equipment and materials evaluation, optimal system programming, production process monitoring, and color reproduction quality research. In many cases, however, test images are poorly designed or improperly chosen, thus distorting the evaluation process. Guidelines are presented in this paper for the design of both analytical and pictorial test images for purposes of color image technology evaluation. A quasi-Mondrian image design strategy is suggested for exploratory color reproduction quality research purposes.

References

1. Field, Gary G. and George W. Jorgensen, *Test Images for Printing*, Graphic Arts Technical Foundation, Pittsburgh, PA, Second Edition, 1984.
2. Jorgensen, George W., "Preferred Tone Reproduction for Black and White Halftones," in *14th IARIGAI Conference Proceedings* (ed., W. H. Banks), Pentech Press, Plymouth, Devon, pp. 109-142, 1977.
3. McCamy, C. S., H. Marcus, and J. G. Davidson, A Color Rendition Chart, *Journal of Applied Photographic Engineering*, vol. **11**, pp. 95-99 (1976).
4. Field, Gary G., *Color and Its Reproduction*, Graphic Arts Technical Foundation, Pittsburgh, PA, second edition, pp. 247-283, 1999.
5. McCann, John J., Magnitude of Color Shifts from Average Quanta Catch Adaptation, *IS&T/SID Fifth Color Imaging Conference Proceedings*, pp. 215-22 (1997).
6. Pointer, Michael R. and Robert W. G. Hunt, A Color Reproduction Index, *IS&T/SID Second Color Imaging Conference Proceedings*, pp. 180-182 (1994).