Using Drum and Flatbed Scanners for Color Image Quality Measurements

D. René Rasmussen, Bimal Mishra and Michael Mongeon
Xerox Corporation, Webster, New York

Abstract

Many image quality measurements of color printer output, that have traditionally been performed using expensive scanning microdensitometers can today be replaced by measurements based on digital images recorded with drum or flatbed scanners. This presentation will discuss reproducibility and accuracy of such measurements for several different image quality metrics, and will also discuss how the design of the image quality metric is important in order to obtain a high level of consistency.

Introduction

Manufacturers of marking systems, be it offset presses, laser printers, ink jet printers, proofing systems or any other marking technology, need methods to evaluate the quality of print samples. Different methodologies exist to address this need, here broadly categorized as Preference Evaluation, Attribute Evaluation, and Metric Evaluation. These categories, and in particular the Attribute Evaluation method, have been described elsewhere.[1]

The focus of this paper is instrumented metrics for evaluation of print quality (subsequently referred to simply as "metrics"). While in the past image quality analysis was nearly impossible without expensive data acquisition instrumentation and special purpose software, one can today purchase relatively inexpensive CCD cameras or scanners as well as software packages which can easily be applied to provide some sort of "image quality metrics". There are many different types of application of such metrics, from benchmarking activities, through various stages of product development, to manufacturing and print quality assurance, and each application has its own set of requirements to the metrics and analysis system.[2]

Given such different applications and requirements, one approach—not advocated here—would be to use different and incompatible analysis systems and metrics for each application. However, it is an important point of this paper, that the advantages of fast and inexpensive image capture devices, can be realized without such large compromises with respect to measurement standards. Analysis systems and metrics can be designed such, that results obtained in one phase of product development, can much more easily be directly applied in the next. For example, "technology-independent appearance metrics"[2] that are necessary for benchmarking purposes and perhaps were evaluated using highly accurate drum scanners, can continue to be used for fast turn-around analysis in fixture labs using inexpensive, and often slightly less accurate, flatbed scanners.

Such versatility does not come for free, of course. The analysis system and metrics must be carefully designed with that in mind. It has been demonstrated that when standard metrics are analyzed with a single drum scanner using sampling resolutions in the range 400-4000dpi, the results can vary significantly.[3] Further variation can be expected when such metrics are evaluated with different scanners which differ in resolving power (even at the same sampling resolution), as well as in other characteristics that affect the scanned image.

This paper addresses possibilities and limitations with respect to defining standard, device-independent metrics. In particular it examines reproducibility both between different types of scanning devices and among a large set of similar scanning devices. Even more important than reproducibility, is the question of to what degree an appearance metric correlates with human perception of quality, however, that issue will not be addressed in this paper.

Image Analysis System

This section briefly describes the image analysis system and scanners used for the experiments reported in this paper. The image analysis software system, called IQAF, was developed within Xerox Corporation, initially to allow the application of standard metrics to simulated, digital images obtained from mathematical models of marking engines. The system is now widely used across Xerox for measurements on hardcopy samples, in dedicated image quality measurement labs and by product development groups.

In the past, Xerox has used a one-of-a-kind scanning micro densitometer as reference instrument, to define standards internally to Xerox. However, for many new metrics, scanning micro densitometry is not a feasible option, even as a reference instrument. Therefore, drum scanners have been introduced to take the place of reference instruments, and whenever possible measurements on the reference drum scanner are traceable back to measurements performed with the scanning micro densitometer.

Scanners

The scanners considered here were not designed to be used for image quality measurements, but were designed for the graphic arts industry. This means that the RGB output of the scanner often is not specified in well-defined physical terms, and the characteristics of the scanner (e.g., aperture size and shape) is often not disclosed by the scanner manufacturer, much less specified. Table 1 shows the scanners typically used with the IQAF system. The scanners made by ScanView are expensive, high-end scanners, used by dedicated IQ measurement labs, while the Umax scanners are relatively inexpensive. This table is not meant to imply that the only or most significant scanner characteristic is the so-called "resolution"—there are other characteristics that are equally or more important.

The "true optical resolution" typically advertised by scanner manufacturers has only little to do with the ability of the scanner to resolve fine details, expressed for example by the modulation transfer function. It refers normally to the sampling resolution, regardless of whether the optical properties of the scanner allows details on that spatial scale to be resolved. (The high sampling resolutions are useful in the graphics industry as an efficient means of enlarging an image).

The drum scanners have a number of advantages over the flatbed scanners. Firstly, the motion quality of the drum scanners is far superior to that of the Umax scanner we have tested (we have not tested motion quality of the F8 flatbed). The drum scanners use photomultiplier tubes as sensors (one for each R/G/B channel), which have excellent noise characteristics and furthermore avoid the cell-to-cell variation of a scan array. The drum scanners use fiber optics to illuminate a small spot on the sample, and has a geometry close to 45° (illumination) - 0° (sensor) which is a commonly used standard. Contrary to this, the flatbed scanners use fluorescent tubes which illuminate a large part of the image at a less well-defined geometry. As a result of the illumination system in flatbed scanners, they are more susceptible to problems caused by integrating cavity effect, that is, due to light reflected from the print being scanned and subsequent internal reflections in the scanner, the effective illumination intensity of the print sample depends on the image on the sample, and thus can vary across the page.

The main advantage of the flatbed scanners lies in the ease of use. The drum scanners have a small depth of focus, and therefore the print sample must be carefully taped to the drum to ensure the sample is absolutely flat on the drum.

Color Calibrations

To calculate an appearance metric from an RGB scan it is imperative that the image is converted into a visual color space such as for example CIELab.[2] The transformation function from scanner RGB to CIELab depends not only on the scanner spectral characteristics (illumination, sensor), but also on the spectral characteristics of the sample being scanned, most significantly on the colorants. This means that for a given RGB scanner it is not possible to have a single transformation to CIELab that works for all types of print samples.

As an example, consider measurements of photopic reflectance of magenta on print samples from an Epson Stylus 800 and from a Xerox DocuPrint C55, using a UMax PLII scanner. The transformations from scanner green response to photopic reflectance for the two print samples differ significantly, with up to 40% difference in slope. That difference can translate into a 40% error in magenta uniformity measurements between the two samples. The significance of using color calibrations for scanner image quality measurements has been demonstrated earlier.[4]

IQAF uses several different types of color calibrations to address these problems. In the most general case, where the measured test element contains more than one separation (C,M,Y, or K) 3x3 color transformations are used, which transform RGB values to one of several visual color spaces. In those cases where the test element is known to contain only one colorant (e.g., Cyan), a "monochrome calibration" that utilizes only one scanner channel (e.g., Red) may be used, for example a 1x1 calibration to calculate L*, or a 1x3 calibration to calculate XYZ. There are trade-offs in whether to use monochrome calibrations, but for example in the case of flatbed scanners where color-color misregistration of the scanner sensors can be significant, monochrome calibrations can improve the accuracy of some measurements.

For a given test pattern, a script is made which specifies the location of all the individual test elements, and which for each test element specifies both the metric to be applied as well as the type of color calibration to be used. A special test pattern, containing 206 different colors, is used to make the color calibrations. This test pattern is first printed under the same conditions as the samples to be measured, then CIELab L*a*b* of the 206 colors are measured with a spectrophotometer, and the samples is scanned under the same conditions as the other test patterns will be. From the CIELab data and the scan file IQAF automatically generates all the necessary calibration files. The accuracy obtained in this way is quite good, typically with an average error of 1-2 ΔE.

Table 1. Scanner Characteristics

Model	Type	Max sampling resolution (dpi)	Illumination
ScanView SM4000	Drum	4000	Tungsten
ScanView SM11000	Drum	11000	Tungsten
ScanView ScanMate F8	Flatbed	4000	Fluorescent
Umax PowerLook II	Flatbed	600x1200	Fluorescent
Umax PowerLook III	Flatbed	1200x2400	Fluorescent

Image Quality Measurements

Prints were made on a Xerox DocuColor 40 electrophotographic printer, using Xerox Color Xpression paper. Table 1 shows the different types of scanners involved in this study. Six pages of test patterns, including the color calibration test pattern, were printed in 20 copies. All copies were scanned on a ScanMate 4000 ("SM" measurements) and on a single Umax Powerlook III ("UM" measurements), both in our lab. 15 copies were then distributed to 15 different labs within Xerox Corporation (referred to as "user systems"), one copy to each lab, and each lab performed measurements with their own scanner. For the following analysis, we will consider separately the group of users with Umax Powerlook (II or III) scanners, and refer to them as the "USER" group. IQAF provides full automation for scanning and analysis with the Umax scanners, so there is very little chance of operator induced errors in the measurements reported here. (Only a subset, 8-10, of the USER data have yet been analyzed for this paper). For the SM measurements a single set of color calibrations were conducted and applied for all the copies, and similarly a single color calibration was applied for all UM measurements. Each user system was color calibrated independently according to the procedure described previously.

Table 2. Scan Parameters

Scanner	Sampling (dpi)	Gamma
SM4000	600	1.8
SM11000	600	1.8
F8	600, 1200	1.8
PL-II	600	3
PL-III	600	3

The scan parameters are shown in Table 2. All scanners were used with 8bits/channel, although they are all capable of higher bit-depth.

Here we will focus on measurements of two metrics: so-called "HDST" and color-color registration. Other metrics that were evaluated are mottle and graniness in solid and tint areas, streaks and bands and line width.

Results

Halftone Depletion Surrounding Text (HDST)

HDST is an Adjacency[1] print quality defect which may be seen near the interface between a high coverage region (e.g. 250% coverage) and a region with much lower coverage. For example, it may be seen as a light halo surrounding solid black text on a halftoned background. Figure 1A shows the analytical test element used to evaluate this print defect. It has 5 horizontal segments with equal amount of C,M,Y from 10% to 50%. Over this background are three vertical line-pairs, also with equal amounts of C,M, Y, and with increasing density. Figure 1B shows a scan of this test element from one of the print samples, where the HDST halo is easily seen.

In addition to the test element shown in Figure 1A, the test pattern contained three more variations of this: one which is rotated by 90 degrees, and two with green background. The four variations are labelled "Ph", "Pv", "Gh", and "Gv".

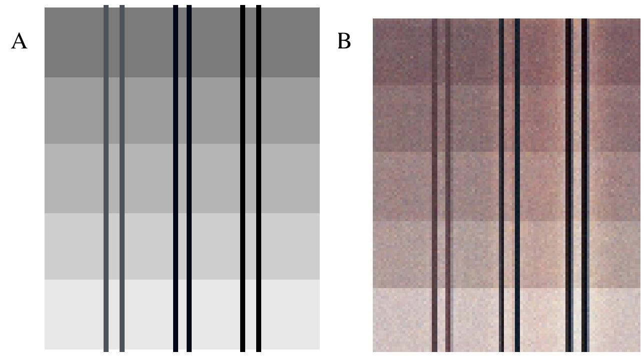

Figure 1. Example of test elements used for HDST evaluation. (A) Digital original; (B) Scan of print sample. The HDST defect is most clearly seen as the light halo surrounding the right most line-pair in (B).

The HDST metric is evaluated independently on each of the 5 background segments. For each segment a background L* level is estimated, which corresponds to the L* which would be obtained if there were no HDST defect. A high-frequency blurring is applied, corresponding to the human visual response at 40cm viewing distance. Then the excess L* above the background level is integrated and taken as a measure of the HDST. Since the image contains more than one separation, all scanner channels are used in the calibration to L*. Figure 2 shows results from the reference drum scanner over all 20 print samples (a total of 20x4x5=400 HDST measurements). At each coverage, the average HDST over the 20 copies is plotted for the four test element variations. At each data point the standard deviation over the 20 copies is shown as an error bar. In Table 3 this is summarized: "Range(SM)" is the range as seen in Figure 2, and "SD(SM)" is the average standard deviation for the 20 data points in Figure 2.

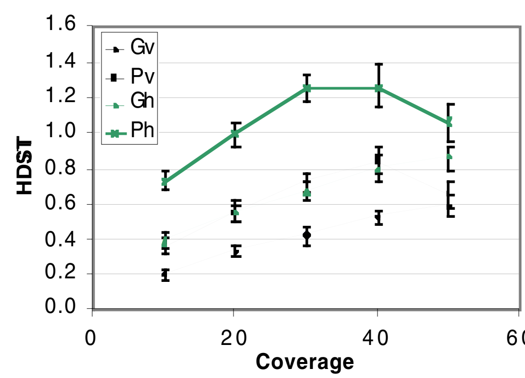

Figure 2. SM measurements of HDST. The error bars show the standard deviation over 20 copies all measured with SM drum scanner.

To examine reproducibility of this measurement with respect to drum scanners as a reference device, Figure 3 shows results the from two drum scanners measuring a single print sample. Not only is there an excellent correlation (R^2=0.98), but the results are nearly identical.

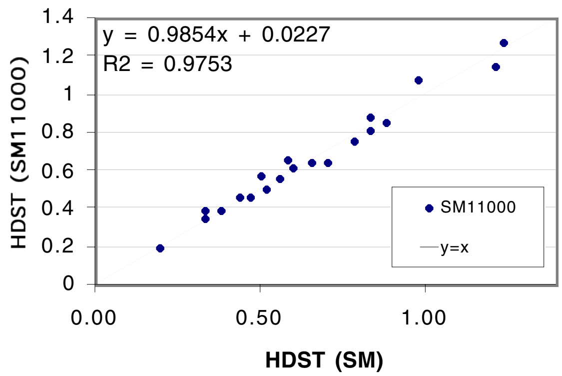

Figure 3 Comparison of measurements performed on different drum scanners.

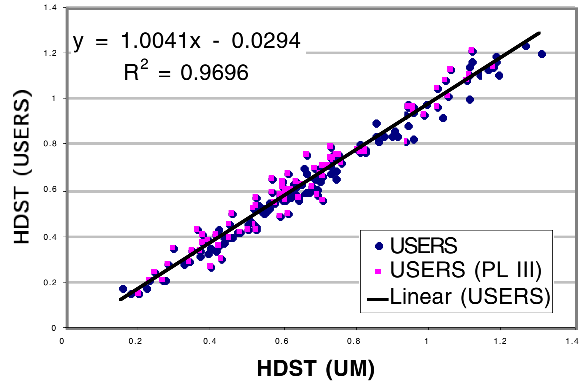

Figure 5 Comparison of measurements on multiple UMAX flatbed scanners (USERS) and single flatbed (UM)

Figure 4 compares measurements from a single Umax PowerLook III (UM) to drum scanner SM, over 15 copies (300 HDST measurements). Again there is a good correlation (R^2=0.94), but here the UM measurements deviate systematically from the SM measurements (roughly 20% smaller). To use the UM scanner for measurements we must assume the regression fit shown in Figure 4, and can then estimate the "true" HDST as defined by the SM. If this procedure is applied to the data in Figure 4, the RMS error between the estimated HDST and the true HDST is 0.08. Such an error in HDST is acceptable for most purposes. In Table 3 "RMS(UM)" is the RMS of the error in estimating SM values from the UM measurements.

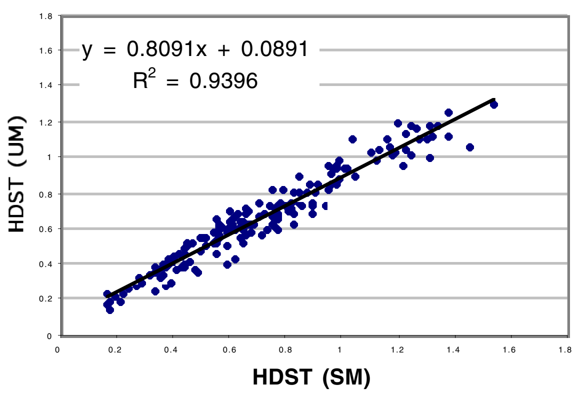

Figure 4 Comparison of measurements with single flatbed (UM) and drum scanner (SM) over 15 copies.

Figure 5 shows measurements on a variety of UMAX flatbed scanners (USER) plotted against measurements on the reference UM flatbed scanner. Those USER data points which come from UMAX PL-III are marked separately, however, there is no statistically significant difference between the two groups of USER measurements. We see an excellent agreement between all the Umax scanners, although with a consistent deviation from the reference values.

In Table 3, "RMS(USER)" is a measure of the accuracy with which SM reference measurements can be estimated from USER measurements, by using the linear regression that was determined from the UM measurements alone (Figure 4).

Color-to-Color Registration

Color registration is measured using a test element as shown (400% magnified) in Figure 6 (and a variation obtained by 90 degree rotation). The results are the offsets between each of C, M, and Y relative to K. This metric does not directly express a human image quality observable, and falls in the category of "diagnostic metrics".[2]

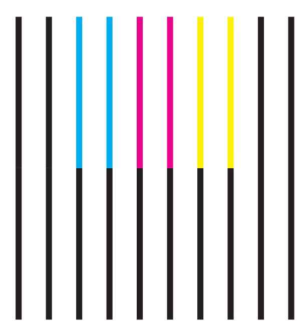

Figure 6 Test element used for measurement of color-to-color registration, shown at 400% magnification. The lines in the lower half are K only. In the upper half, the lines are: 2K, 2C, 2M, 2Y, and 2K.

The metric determines 20 centroid locations in the horizontal direction, 10 of the lines in the upper half of the image, and 10 of the lines in the lower half. By comparison of centroid locations of the C, M, and Y lines in the upper half, to the centroid locations of the corresponding K lines in the lower half, the misregistrations relative to K can be determined. The 2 outer K lines on both sides are used to make corrections for skew introduced during the scanning process. The analysis of each color is done entirely based on a single scanner channel, for example the determination of Cyan to black registration utilizes exclusively the Red scanner channel, and in that way the measurement is not affected by misregistration between the scanner channels. There is no color calibration required for this metric. This approach can determine color misregistration far more accurately than the scanner sampling resolution would seem to indicate.

Each set of test patterns contains 6 instantiations of the test element. Figure 7 shows roughly 50 measurements of C-K misregistrations, measured with the UM and with the SM. The results are nearly identical, and in Table 3 the "RMS(UM)" is a measure of the direct discrepancy between UM and SM, without any correlation correction.

Figure 8 shows a similar comparison between 8 different USER results and the UM results, and again the agreement is excellent.

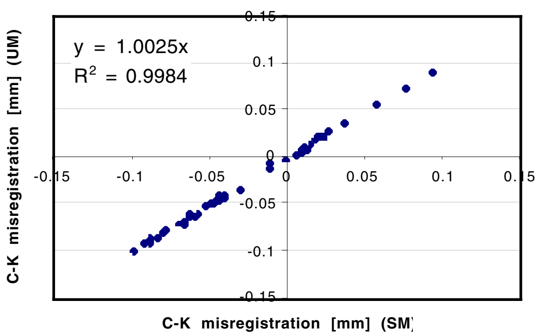

Figure 7. Comparison of cyan color registration measurements on a single flatbed (UM) versus a drum scanner (SM). A total of 50 measurements.

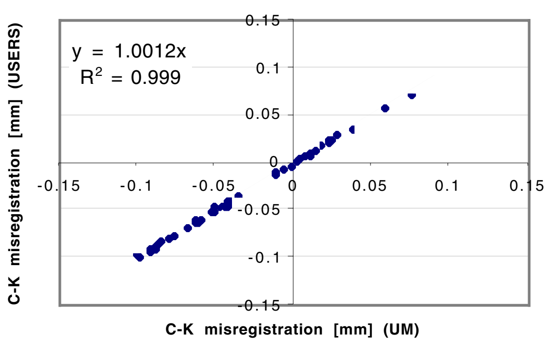

Figure 8. Comparison of cyan color registration measurements on 8 different flatbed scanners (USERS) versus a single reference flatbed scanner (UM). A total of 48 measurements.

Discussion

The results from the two metrics, HDST and color registration, were both encouraging, but different. In both cases the conclusion is that even the relatively inexpensive flatbed scanners used here can provide reliable measurements, but in the case of HDST the measurements are device-dependent, and require a drum scanner to establish a reliable reference measurement system. For color registration measurement it is most plausible that the flatbed scanners will provide accurate measurements even if another printing technology is being evaluated, but for HDST the correlation between UM and SM measurement might depend on the printing technology to some degree, and it is important to use the drum scanner as a reference.

The discrepancy between the HDST measurements is too large to be explained by inaccuracies in the color calibration. A possible cause is the integrating cavity effect (ICE), which affects the UM scans. ICE causes regions in the image which are closer to dark image regions, to be recorded by the scanner as darker than they really are. This means that on the UM scanner, the region surrounding the line-pairs can be expected to be recorded with higher density, than they really have, which would tend to cancel out part of the HDST halo, leading to lower HDST values.

Relatively inexpensive flatbed scanners can be used, but not always without a correlation function (or "fudge factor") if the results are to correspond to standard measurements. An alternative to such "fudge factors" is to perform spatial corrections on the scanned image before further processing, or to otherwise take the device characteristics into account during the evaluation of the metric.

One example of such pre-processing is integrating cavity correction, in which a scanner characterization is performed once, and subsequently is applied as an image-dependent correction before color calibration and metric evaluation. As another example, consider a metric which involves blurring the image according to the human visual perception, before further analysis.[2,5] In this case, the amount of blurring can be adjusted to compensate for the scanner MTF, and thus reduce device-device variations caused by differences in MTF.

Table 3. Overview of Metric Comparison.

Metric	Range (SM)	SD(SM)	RMS (UM)	RMS (USER)
HDST	0.2 to 1.4	0.06	0.08	0.07
Cyan reg.[um]	± 100	N/A	2 (direct)	1 (direct)
Magenta reg.[um]	± 60	N/A	2 (direct)	2 (direct)
Yellow reg.[um]	± 60	N/A	9 (direct)	12 (direct)

Acknowledgements

The authors would like to thank E. Dalal for fruitful discussions, F. Medina for technical assistance, and a number of people at Xerox for providing equipment and data for this study, including S. Reczek, J. Debarr, J. Masseth, M. Rabbani, E. Reavis, K. Stamp, S. Sterling, Q. Yang, B. Zhou, S. Zoltner, and E. Zwartz.

References

1. Edul N. Dalal, D. René Rasmussen, Fumio Nakaya, Peter A. Crean and Masaaki Sato, Evaluating the Overall Image Quality of Hardcopy Output, *Proc. PICS*, pg. 169 (1998).
2. D. René Rasmussen, Peter A. Crean, Fumio Nakaya, Masaaki Sato and Edul N. Dalal, Image Quality Metrics: Applications and Requirements, *Proc. PICS*, pg. 174 (1998).
3. Jim Grice and Jan Allebach, Print quality assessment with a virtual microdensitometer, *Proc IS&T 50th Ann. Conf.*, pg. 446 (1997).
4. William Lim and Suresh Mani, Application of Digital Imaging to Measure Print Quality, *Proc. NIP 14*, pg. 611 (1998).
5. Bimal Mishra and D. René Rasmussen, MicroUniformity: An Image Quality Metric For Measuring Noise, *Proc. PICS*, (2000).

Biographies

René Rasmussen received his Ph.D. in physics from the Niels Bohr Institute, Copenhagen, in 1990. He is Principal Scientist at Xerox Corporation and leads a project on development of Xerox internal standards for color image quality assessment. He has been with Xerox Corporation since 1992, and is a member of the IS&T.

Bimal Mishra is a research member in the Color Science and Image Quality Area at the Wilson Center for Research and Technology in Webster, New York. His current responsibilities include development and implementation of image quality evaluation algorithms. He has a Ph.D. from Columbia University. His thesis was on the application of Brownian motion to visco-elastic liquid simulation. Prior to coming to Xerox, he worked as a visiting researcher at the Courant Institute of Mathematical Sciences in New York.

Michael Mongeon received his B.S. degrees in Mathematics and Mechanical Engineering in 1985 from University of Buffalo and M.S. degree in Imaging Science from Rochester Institute of Technology in 1994. Since 1986, he has worked at Xerox Corporation in Webster, NY, currently in the Wilson Center for Research and Technology. He has experience in color science, image quality, software development, xerographic systems integration, and process controls.