

Image Quality Metrics: Applications and Requirements

D. René Rasmussen,[*] Peter A. Crean,[*] Fumio Nakaya,[†] Masaaki Sato [†] and Edul N. Dalal[*]
[*]Xerox Corporation, Webster, NY, USA
[†]Fuji-Xerox Co., Ltd., Ebina, Japan

Abstract

This paper addresses instrumented metrics for evaluating image quality of hardcopy prints. Image quality metrics can be extremely valuable during all phases of product development, from planning through design and development, to manufacturing. The various applications impose very different requirements on the nature of the metrics and how they can be measured. For example, technology-independent metrics must be used in order to compare different technologies. In this paper the different requirements for image quality metrics and measurements are discussed, using examples from recently developed metrics for uniformity.

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"). Under the best of circumstances we would have industry standards for quantification of all the attributes of print quality, with metrics which could be relied upon by manufacturers to provide guidance throughout the product development cycle, and which could be used to convey more useful information to the customers than the statements commonly used today, such as "photographic quality" or "1200dpi quality". Most people who work on image quality related issues would quickly agree with this statement, however, each one having their own specific application and requirements in mind—in fact many standards do exist for print quality metrics, but very few are applicable beyond a narrow scope limited either by the technologies they apply to or by the phase in the product development cycle where they are useful. For example, the grain index used in the photographic film industry does not work well on printed images, and although a simple line width and density measurement technique can be used across different technologies for engineering applications, it does not adequately define customer perceptions of lines to the point where it is a sufficient metric for other applications.

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 yield "image quality metrics". However, such systems can be extremely limited in applicability and can easily provide misleading information if used outside their intended scope.

On the other hand, the widely available image capture systems, combined with the computational power of typical PCs, open the possibility for creating metric standards with far greater applicability than in the past, and it is worth while to consider carefully the minimum requirements to such standards. This paper discusses the different applications of image quality metrics and the minimum requirements they each impose on the nature and implementation of metrics. The concepts are briefly illustrated by an example of metrics for an image quality problem known as banding.

Metrics: Applications And Requirements

The image quality metrics discussed in this paper are those that address structure in the printed image, whether the structure be intentional or not. Examples are metrics to quantify non-uniformities (e.g., graininess) in areas of the image which are intended to be uniform, or metrics to quantify the extent to which the edge of a line is straight (e.g., free from raggedness and jaggies). This type of metrics has traditionally been based on reflectance profiles recorded with scanning microdensitometers. There are two fundamentally different types of print quality metrics:

(1) *Appearance metrics* which are used for quantification or prediction of how an attribute of image quality will be perceived by the customers. The metric has a monotonic relationship with the quality as perceived by the customer.

(2) *Diagnostic metrics* which provide accurate assessments of the physical characteristics of the image. In this

case there may or may not be a direct relation to customer perceptions of quality.

As an example, consider a metric for uniformity of halftone patches from an ink jet printer, defined simply in terms of dot size and placement accuracy. Such a diagnostic metric could not directly be used to predict the perceived uniformity of printers using different inks of different optical density. Table 1 gives an overview of some of the important applications of and distinctions between the two types of metrics, and the details are discussed in the remainder of this paper.

Table 1. Appearance metrics versus diagnostic metrics.

	Appearance metrics	Diagnostic metrics
Purpose	• Predict customers perceptions of image quality attributes. • Help assess acceptability of a product.	• Quantify different levels of print quality with high resolution, for feedback to engineers. • Help reduce or eliminate a print quality problem.
Test patterns	• Analytical test patterns which facilitate measurements. • Should not unrealistically stress the printer. • Example: Use native halftone screen.	• Should maximize sensitivity of measurements. • Example: Use halftone screen that stresses photoreceptor uniformity.
Data acquisition	• Typically use data representative of human visual system. (e.g., photopic reflectance or CIELab)	• Use filters to maximize signal-to-noise ratio (e.g., R/G/B filters for C/M/Y).
Metric algorithm	• Must directly incorporate human visual system factors to be robust for all technologies.	• Should provide engineering detail (e.g., frequency of bands).
Validation	• Must be validated across many technologies.	• Often limited to single technology or product family.

Image quality metrics are used for many different purposes throughout the product development cycle. To *set product specifications* via benchmarking of competitive products it is often not possible to use diagnostic metrics especially in the case where the technologies differ so that the physical characteristics of the images cannot be directly compared, see for example the comparison of graininess measures for silver halide and electrophotographic prints by Bouk and Burningham.[2] For this phase, *metrics that directly relate to appearance are required*, relatively few measurements are performed and the emphasis is on accuracy rather than throughput.

During the second phase where fixtures are being tested and the *product design* optimized, the requirements are different. There is a need for immediate feedback in terms of the impact of design parameters on the image quality, however, the trends are often more important than the absolute values. In addition to the metrics used for the specifications, the engineers need feedback which can help diagnose and solve the problems. Often computer modeling is used to help optimize the design, for example for halftone screen design, and for a metric to be applicable for that purpose, it must be independent of the image capture hardware, that is, it must be defined in terms of the physical image, deconvoluted from any image capture characteristics.

Finally, for manufacturing quality control the throughput is of highest priority. Typically, a relatively small set of image parameters have been defined which need to be monitored, and which could be monitored with metrics other than those used in the product specifications. However, today many appearance metrics can be evaluated with simple and fast image capture systems using scanners or CCD cameras, thus eliminating the need to translate product specifications in terms of metrics that are specifically developed for manufacturing purposes.

Appearance Metrics and Technology Independence

Why we need technology-independent metrics and test patterns

As pointed out above, appearance metrics are not necessary for all applications, yet we argue here that in the long run it is well worth the investment to establish appearance metrics rather than rely on diagnostic metrics. From the point of view of setting industry standards, technology-independent metrics must be used, and the best candidate for that is via metrics that are based on appearance. Even for a given product program there are strong advantages to the use of appearance metrics. The alternative to appearance metrics is to establish correlations directly between diagnostic metrics or even technology variables and customer preference[3]. Such correlations are not likely to hold when other technology parameters are changed, and thus they may have to be re-established repeatedly during product development.

To set specifications and test performance against the competition with appearance metrics, it is imperative to use "technology-neutral test patterns". While highly specialized and stressful test patterns are ideal to diagnose certain technical problems, they do not represent the real customer documents, and therefore are of little value for technology comparison. Today's printers are complex systems with subsystems that are cross-optimized to yield the best overall system performance and therefore, there is only limited interest from a benchmarking point of view to quantify the quality of each subsystem using diagnostic test patterns and metrics.

The implications on the test pattern design are far reaching. This is especially so, because today's printers rely heavily on image processing during the image path before the marking engine is reached. Images are segmented and processed independently, which means, for example, that the quality measured on a graphics line in the test pattern may not correspond to the quality found on text. The challenge is thus to make test patterns that follow the correct image path which allows the printer system (including OS and printer driver) to perform at its best, yet are still amenable to instrumented analysis.

Implications for metric data acquisition and algorithms

A prerequisite for an appearance metric is image data in well-defined physical terms that can be related to how the human visual system perceives images. For color accuracy metrics it is well appreciated that colorimeters or spectrophotometers are necessary instruments, however, when it comes to color image structure analysis it is unfortunately not unusual to see the analysis performed in terms of uncalibrated RGB data.

The ideal image capture apparatus for appearance metrics is the human visual system [HVS] —except that the data are not readily available for computer processing. The second best is an image capture device which is "better" than the HVS and is calibrated and characterized well enough that the excess performance of the device relative to the HVS can be artificially "blurred away". Many scanners and CCD cameras can be used in this manner. Figure 1 illustrates the processing concept: The physical image recorded by the capture device and calibrated in the proper color space is passed on to a model of the HVS (with the device function factored out if necessary) yielding a representation of the perceived image. Appearance metrics can now be calculated from the perceived image, if necessary by utilizing a similarly constructed "perceived perfect image" to calculate visual differences.The simple diagnostic metrics would typically calculate metrics directly from an uncalibrated, physical image. Needless to say, the HVS model is the weak link in this chain, however, even our current limited understanding of the HVS offers a definite advantage to metrics that utilize this approach.

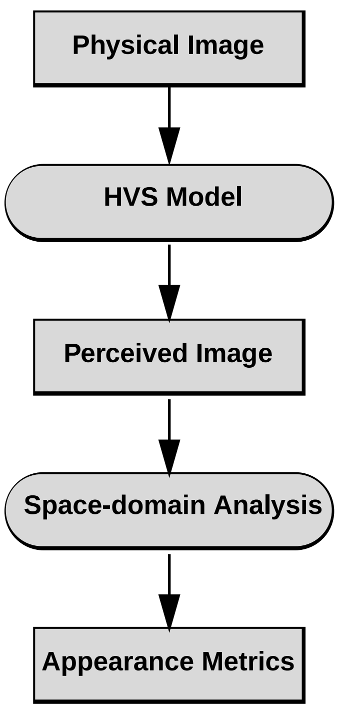

Figure 1. Schematic of the processing behind an appearance metric.

Example: Metric for Quantification of Banding

To illustrate the concepts of appearance versus diagnostic metrics discussed above we will consider a metric for banding. Banding refers to macro-uniformity problems manifested as color variations (or density variations in the case of black and white prints) extended in one direction typically either parallel or perpendicular to the paper feed direction. Banding can be observed on prints from practically any technology, for example as a result of print head signatures from ink jet printers, as a result of donor roll eccentricities for electrophotographic printers, or as a result of motion quality problems. The profiles in Figure 2 show the variation in photopic reflectance relative to the average across prints from several different technologies. The test pattern was a uniform 30% gray page, and the profiles show a 60mm region representative of the entire page. The first point to notice is that a metric based on RMS of L^* from the average L^*, will not work. The reason is that the HVS is more sensitive to variations that occur over short distances as to gradual variations over long distances. This is easily understood from the human visual contrast sensitivity function, see for example Cornsweet.[4] When judging the 3 print samples visually, the ink jet sample is characterized by a periodic pattern corresponding well to the thin light bands of around 11mm period seen in Figure 2A. The electrophotographic sample is visually characterized by several weak bands and a single more objectionable band at the location indicated by the arrow in Figure 2B. The dye diffusion sample is visually characterized by a few distinct, relatively broad bands.

From a diagnostic point of view each of these prints could be analyzed by separate techniques, for example the ink jet prints could be analyzed in terms of misdirected jets, and the motion quality related bands could be analyzed in the frequency domain with amplitude spectra to provide information on the constituent frequencies. These methods alone, however, cannot be used for a general metric to predict the human perception of the severity of the banding problem. For example, although the amplitude spectrum is an important characteristic which could be used in combination with the visual sensitivity function to predict the visibility of sinusoidal patterns, it fails to characterize the visibility of non-periodic patterns. This is illustrated in Figure 3, which shows synthetically generated vertical bands. These images are not related to the print samples used for Figure 2. The bands shown in Figure 3A and 3B have identical amplitude spectra, yet the single band in 3B would be

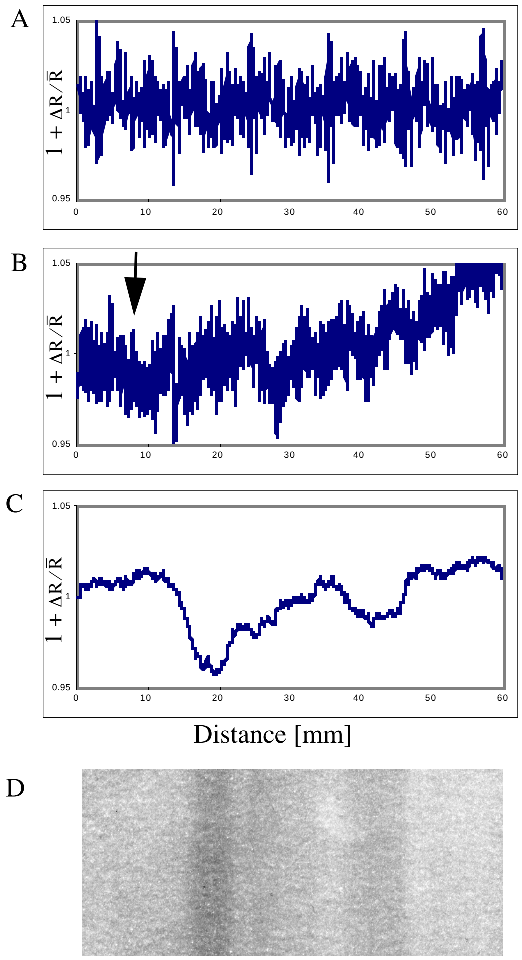

Figure 2. Profiles showing variations in reflectance across bands from prints made with ink jet (A), electrophotography (B), and dye diffusion (C) technologies. (D) shows the image corresponding to the graph in (C), although with increased contrast to assure reproduction.

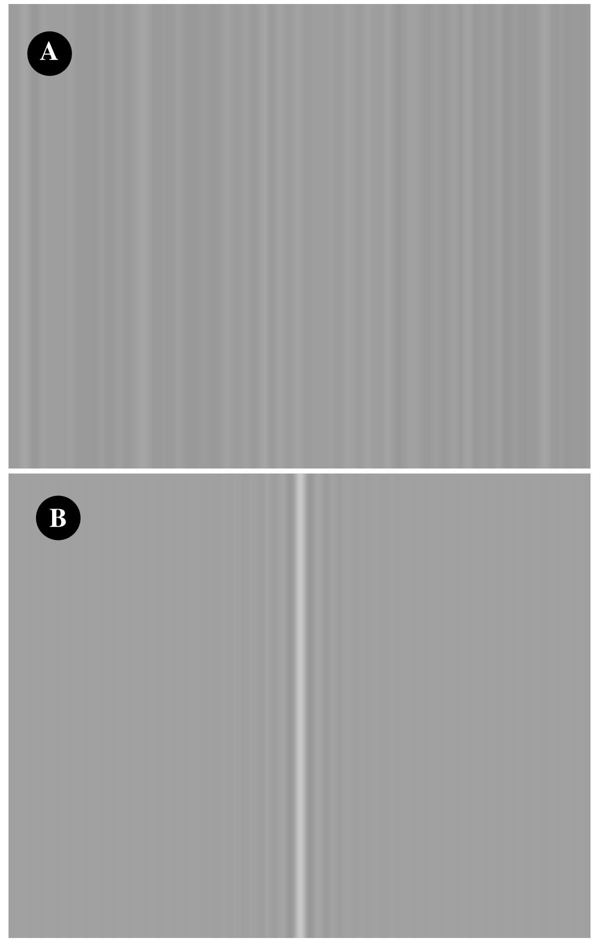

Figure 3. The bands in (A) and (B) have identical amplitude spectra. The difference lies only in the phase relationship between the harmonic components. The bands in (A) can be at or below perception threshold while the band in (B) is highly objectionable. Here, all the amplitudes have been increased in order to assure reproduction.

far more objectionable than the many bands in 3A which are close to or below perception threshold. Therefore, any metric based solely on the amplitude spectrum is bound to fail as a general metric for banding.

Frequency-Corrected Space-Domain Analysis

Following the approach outlined in Figure 1 with a simple model of the HVS, the profiles in Figure 2 can be analyzed in the space domain. Figure 4 shows the same profiles after the HVS model has been applied. These profiles represent more accurately what we perceive when looking at the prints than do those in Figure 2. For example, the single most objectionable band which is hardly detectable from Figure 2B, is easily identified in Figure 4B.

In our recent work, metrics based on analysis in the space domain have proven capable of predicting the overall objectionability of the bands on print samples from ink jet, electrophotography, lithography and dye diffusion. The space-domain analysis approach is very powerful, and has been applied to other image quality attributes, for example for analysis of halftone texture using the S–CIELab color difference metric.[5]

Summary

To effectively communicate and compare image quality capabilities across different technologies either within a corporation or at an industrial level, it is necessary to use either high-level attributes,[1] or technology independent appearance metrics. Although many image quality analysis systems and metrics are being offered as providing "standard" solutions, they are often limited in applicability, and may not be able to serve even a specific product throughout the development cycle. The minimum set of requirements for appearance metrics and test patterns has been discussed, and an approach for development of appearance metrics was outlined. Given the state of image capture hardware and data processing capabilities, the pursuit of appearance metrics is worthwhile the effort, although much research into human visual perception, both perception thresholds and especially perception above the threshold, is required.

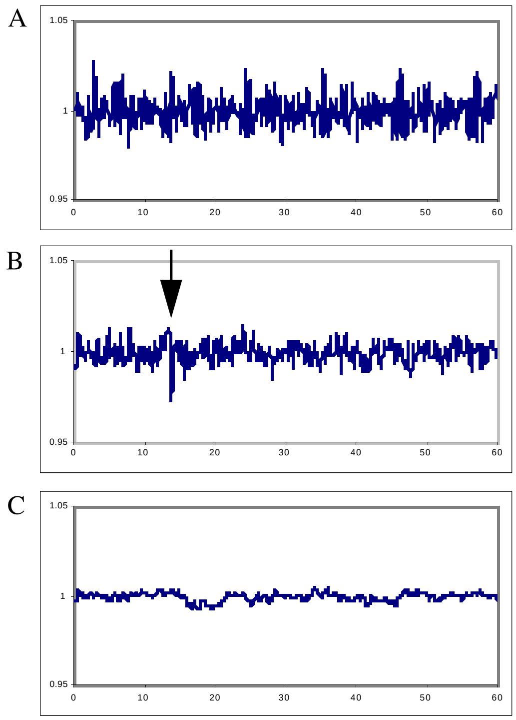

Figure 4. Profiles corresponding to Figure 2 after a simple HVS model has been applied. The significant difference is that low and high frequency variations have been damped, leading effectively to an enhancement of abrupt density variations.

References

1. E. N. Dalal, D. R. Rasmussen, F. Nakaya, P. A. Crean and M. Sato, Proc. IS&T 1st PIC Conference, Portland (1998).
2. T. Bouk and N. Burningham, Proc. IS&T 8th Int. Cong. on Adv. in Non-Impact Printing Tech., 506-509 (1992).
3. P. Engeldrum, J. Imaging Sci. and Tech., **39**, 312-318 (1995).
4. T. Cornsweet, "Visual Perception," Academic Press (1970).
5. X. Zhang, D. A. Silverstein, J. E. Farrell and B. A. Wandell, IEEE Compcon (1997)