Evaluating the Overall Image Quality of Hardcopy Output

Edul N. Dalal, D. René Rasmussen,* Fumio Nakaya,†*
*Peter A. Crean * and Masaaki Sato †*
** Xerox Corporation, Webster, NY 14580*
† Fuji-Xerox Co., Ltd., Ebina, Japan

Abstract

This paper addresses the need for a manageable method of evaluating the overall image quality of hardcopy output from printer systems so that the systems can be compared in a meaningful way. Image quality metrics are widely used and are very objective, but are unsuitable for this purpose because there is no good way of extracting overall image quality from the large number of possible metrics. On the other hand, preference measurements do provide information on overall image quality in a customer-relevant form but are notoriously subjective. Moreover, it is difficult to extract from them information needed for engineering improvements, since preference can be severely impacted by poor performance on a single attribute. In this paper, a comprehensive high-level set of image quality attributes is introduced for the purpose of evaluating overall image quality. An example of such a set of attributes is proposed and discussed.

Introduction

Image quality is the overall measure of success of a color printing system. It is an important customer requirement, along with other requirements such as cost, productivity, connectivity, and reliability. Owners of office printing systems typically rate image quality as the number one reason for selecting the product.

This paper deals with the image quality of color documents. Image quality has traditionally been evaluated in two very different ways: *image quality metrics*, which are objective measurements of the physical characteristics of an image, and *image preference*, which is an overall measure of how well customers like a given image. These aspects of image quality are discussed, along with their strengths and weaknesses. Then the concept of a set of *image quality attributes* is introduced, which enables evaluation of *overall* image quality.

Image Quality Metrics

Image quality metrics are well defined procedures for quantitatively measuring specific image quality features. Each procedure is associated with only certain elements of an image, typically something easy to measure physically, like line density.

Not all metrics take human visual perception into account. Line density, for example, is a purely physical measurement. In order for the output of a metric to be proportional to the human response, however, the functions of the human visual system must be factored into the image quality metric. Studies on human perception have been published extensively, e.g., by Cornsweet[1] and by Wandell,[2] but key perception data are still lacking. For instance, in quantifying the perceptibility of color variations, the spatial frequency dependence of lightness L^* is known, but not of chroma C^* or hue h^* for base colors other than neutral.

Rasmussen et al[3] discuss the importance of perception-based image quality metrics. IQAF (Image Quality Analysis Facility) is a proprietary software package for determining the image quality of hardcopy output from printing systems, using data captured from a variety of input devices. It is widely used internally at Xerox (Xerox Corp. and Fuji Xerox Co., Ltd.).

Image Preference

Image quality metrics described in the previous section provide much useful engineering information, but do not directly tell us how well a customer would like a particular image. This is known as image preference, and can be determined only by showing customers (or surrogate customers) the image and getting their response. For this technique to be effective, "test targets" or analytical images are not suitable. One must use actual customer-like images, such as pictorials, graphics, newsletters, etc. Even then, image preference can be difficult to quantify because it is so subjective. Typically large numbers of customers must be surveyed in order to get reliable data. We are currently developing an image preference system for color printers, called

QPS (Quantitative Preference System). For each image, a fixed preference scale is established using anchor prints. When asked to judge image quality, customers place a print according to their preference on the pre-defined scale.

A number of requirements must be met for a quantitative preference system to be successful. First, there must be agreement on the standard digital image files (DIFs). All organizations using such a system must use the same DIFs for standardization in measuring image preference, because preference data obtained on different images cannot be compared in a quantitative manner. These DIFs must cover all major types of customer documents, which includes pictures, graphics and newsletters. A preference scale must be developed for each DIF. Each preference scale must be duplicated and made available to relevant users, and suitably archived to avoid deterioration. The number of DIFs must be kept to a minimum because of the significant effort involved in developing, duplicating and archiving the scale for each DIF.

A major issue in using image preference as a measure of overall image quality is that color rendering (i.e., how each color in the image is interpreted by the device) is critical; small variations in color can in many cases have a profound effect on image preference. Successful color rendering is often more art than science, and is in general not an intrinsic property of the marking technology. In addition, if one or two aspects of image quality are poor, image preference could be low and insensitive to good performance in other aspects. These issues need to be comprehended in designing and using a preference system.

Inherent Drawbacks Associated with Metrics and Preference

Image quality metrics, described earlier, are objective, quantitative, insensitive to color rendering issues, and can examine specific aspects of image quality regardless of deficiencies in other aspects. But they are ill suited as high-level image quality descriptors, because it is not feasible to combine the large number of disparate metrics, which are far from independent, into a concise view of overall image quality. There are several dozen image quality metrics in use at Xerox, and there would be many more if procedures existed for everything that the engineers would like to measure. Moreover, some metrics can produce an infinite number of output values; e.g., the metric for graininess of halftones produces results that vary with image color, halftone coverage, etc. If one wished to compare the overall image quality of two output devices using image quality metrics, one would need to compare literally thousands of numerical values. This is difficult to do in a meaningful manner.

As discussed above, image preference does not have these drawbacks, and could be used as a measure of overall image quality. However, it is subjective, intrinsically qualitative (difficult to quantify) and strongly dependent on image content. Moreover, the influence of color rendering on preference is a major concern, together with a high sensitivity to the weakest link in the overall image quality. For these reasons image preference is often not a good way of measuring overall image quality, particularly for immature technologies, where some aspects of image quality still need to be worked on, and where the color rendering solutions have not yet been adequately addressed.

Because of these inherent drawbacks associated with image quality metrics and image preference, image quality attributes have been introduced as another way of representing the overall image quality of a printer system.

Image Quality Attributes

An image quality attribute is a high-level image quality descriptor, such as Line Quality, which describes the overall quality of lines in a given output. What is meant by "high-level" in this context is discussed below, and standard image quality attributes are listed later in this section. Several image quality metrics may be associated with a given attribute: e.g., metrics such as line density, line width, line raggedness, etc. are used to measure different aspects of the Line Quality attribute.

A high-level image quality descriptor is one which is suitable for concisely describing the overall image quality of a given device or technology. This requires that the overall image quality can be described with relatively few such descriptors, and that they be essentially orthogonal to, or at least independent of, each other. At the same time, the set of attributes should be comprehensive, i.e., should cover all potential image quality issues for a very wide range of applications. Typical uses for high-level image quality descriptors range from image quality benchmarking, to support for funding decisions for competing technologies.

Image quality attributes combine some of the best features of both image preference and image quality metrics. They have some of the objective and quantitative nature of metrics. Yet the overall image quality of a given output device can be meaningfully represented with a relatively small number of attributes. A drawback is that at present many image quality attributes can only be evaluated visually, although there are plans to enable their instrumental measurement in the future.

A long-term goal is to be able to predict image preference from measurements on analytical images, using statistical correlations. This is not feasible to do directly with metrics because of the large number of metrics one would need to consider. Image quality attributes provide a bridge between image preference and image quality metrics, since image quality attributes can be correlated with both of them.

The DAC System of Image Quality Attributes

Image quality attributes used at Xerox are those specified by the DAC (Document Appearance Characterization) system. The original DAC system was created by a committee led by R. Gruber in 1992-93, to evaluate and compare image quality characteristics of competing technologies. Significant changes have been made since then. Figures 1 and 2 show the current DAC attributes in the form of two radar charts, which is how they are usually presented.
The DAC attributes are rated on an absolute scale of 0 to 100, with 100 indicating perfection.

The DAC system includes a specification of a standard set of digital test images and a detailed procedure for evaluation of a printer system on the basis of these images. For most attributes, the evaluations are performed visually by an expert panel, but the standard procedure makes the process fairly objective, to the point where good agreement is obtained between evaluations done independently by different groups.

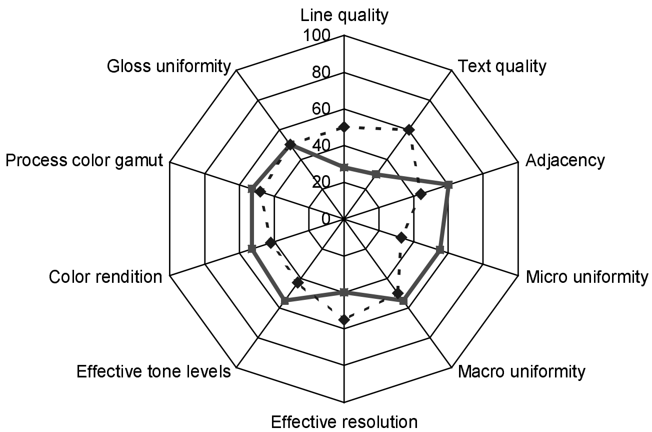

Figure 1: The DAC image quality attributes: radar chart #1, Basic Image Quality attributes. The solid and dotted lines illustrate the image quality of two hypothetical printers.

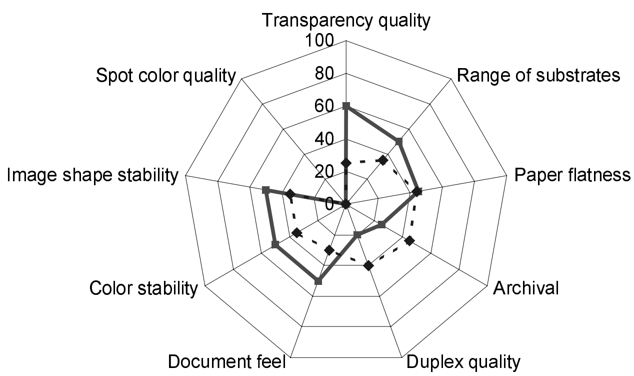

Figure 2: The DAC image quality attributes: radar chart #2, Materials and Stability attributes. The solid and dotted lines illustrate the image quality of two hypothetical printers.

The criterion used in defining the DAC attributes is that they should be appearance-based, not technology-based. The technological origin is irrelevant for DAC rating, and one technological problem may affect several attributes. For example, color-color registration is not considered to be an attribute, because poor registration is not itself an appearance issue but will cause shortfalls in other DAC attributes.

Each DAC attribute has an associated set of metrics and technological issues. For example, the metrics associated with the Micro-uniformity attribute include solid and halftone graininess and visible halftone screening. The technological issues associated with this attribute include process noise, stochastic (error diffusion) screening, and coarse screening.

The Basic Image Quality Attributes are collected in the first radar chart. These are visually relevant attributes that can be evaluated from a single set of analytical images on the standard or default substrate.

The Line Quality attribute refers to the overall quality of lines in the images. It takes into account problems such as: jagged lines due to low printer resolution; fuzzy or ragged lines due to ink bleed, toner splatter or poor registration; lines with inadequate density; and lines with poor discriminability due to severe width quantization. Similarly, the Text Quality attribute refers to the overall quality of text. This is influenced by all the factors relevant to Line Quality, together with issues such as reproduction of serifs and line placement accuracy.

The Adjacency attribute takes into account any defects associated with edges between two colors (including white). This includes problems such as trail-edge deletion, intercolor bleed, edge enhancement, etc. The Micro-uniformity and Macro-uniformity attributes deal with non-uniformity in areas that are intended to be smooth and uniform. Micro-uniformity is restricted to problems that are visible in small areas (defined as an aperture of 6mm diameter). On the other hand, problems with Macro-uniformity are most visible in large uniform areas.

Effective Resolution is related to pictorial sharpness, and refers to the ability to distinguish fine detail, especially at low contrast. It is related to but distinct from print engine resolution and addressability. The Effective Tone Levels attribute refers to the quality of single-hue sweeps, including freedom from contouring, and the ability to discriminate density levels at very low and very high density levels.

The Color Rendition attribute deals with color management issues, including gamut mapping and smoothness of multi-hue sweeps. Process Color Gamut is directly related to the range of colors printable on a process color system. It is a non-linear function of the gamut volume, or the number of Pantone colors included in the gamut.[4] The Gloss Uniformity attribute refers to the uniformity of the glossy or specular component of the light reflected off the image. It includes gloss variations within a nominally uniform area (micro-gloss) as well as gloss differences between bare paper and image areas of various density levels.

The Materials and Stability Attributes are collected in the second radar chart. These attributes depend significantly on marking materials and substrates: Transparency Quality, Paper Flatness, Archival Properties, and Document Feel. Range of Substrates refers to the variety of papers and other substrates on which image quality comparable to the standard substrate can be achieved. Duplex Quality and Spot Color Quality refer to image quality achievable with these features. Color Stability is a measure of repeatability over time, environmental conditions and machine-to-machine.

Relationship between Metrics, Attributes and Preference

Image quality metrics, image quality attributes and image preference have been discussed in the preceding sections. Their characteristics are compared in Figure 3. Image quality metrics are intended for engineering and design applications, while image preference is suitable for customer-related applications such as marketing. Image quality attributes are ideal for planning and other high-level applications, but could have some utility in engineering and marketing as well.

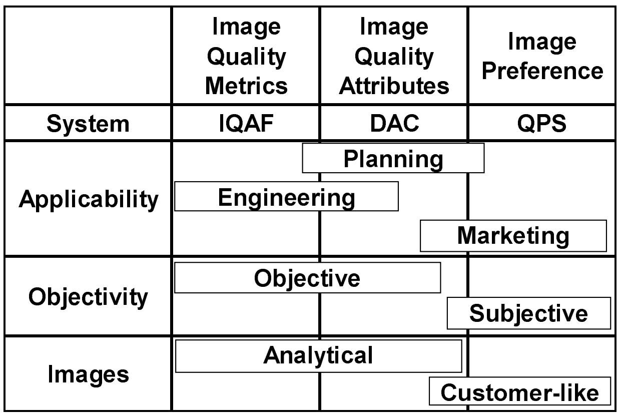

Figure 3: Features of the three domains of image quality.

The application of image quality in industry can be described in terms of the Image Quality Circle™ (Engeldrum[5]) shown in Figure 4. At the top of the circle is Customer Quality Preference. Evaluating this is the ultimate goal of any image quality activity. In order to optimize the performance of a color printing system, engineers need to know how changes in the technology variables (e.g., bias voltage, roll spacing, etc.) affect customer quality preference. Attempts are often made to develop a direct correlation between customer quality preference and technology variables, as shown by the heavy arrow in Figure 4. It is possible to do this by printing images at various settings of the technology variables and having customers judge the quality of the output. But whenever any changes are made, the correlation has to be reconstructed because the previous one is no longer valid.

A much more robust correlation can be developed by taking the indirect route indicated by the smaller arrows in Figure 4. This consists of a series of definable and measurable steps. Thus technology variables can be related to physical image parameters, preferably in terms of physically meaningful system models. The physical image parameters can be related to customer perceptions by visual algorithms, and these in turn can be linked to customer quality preference by image quality models. When changes are made to the system, this approach can handle them with little modification because the different stages are separated, and the relationships between them are defined in terms of extendible models. Although the indirect route is more complicated and takes longer to establish, in the long term this approach is clearly superior.

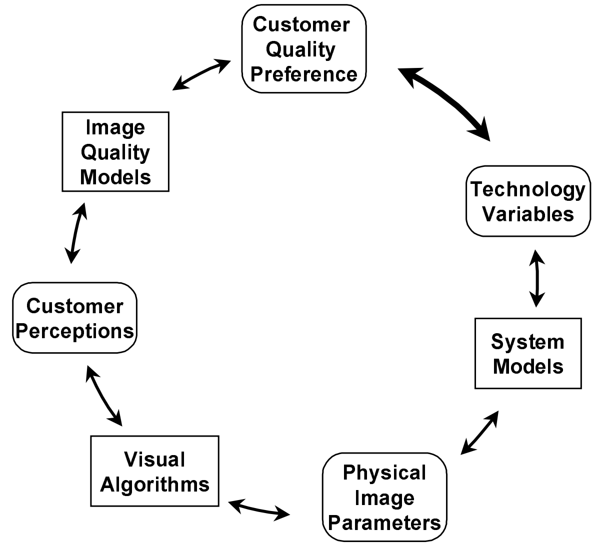

Figure 4: Image quality circle™ (Engeldrum[5])

A similar path is shown in Figure 5, where the technology variables are linked to image preference *via* image quality metrics and image quality attributes. As in the previous figure, the heavy arrows refer to direct determination of preference from the technology variables. The lighter arrows indicate the preferred analytical approach, in which image quality metrics are determined from analytical images. Ideally, DAC attributes would be calculated from the metrics, though the attributes can also be determined visually from analytical images, as shown by the dotted line.

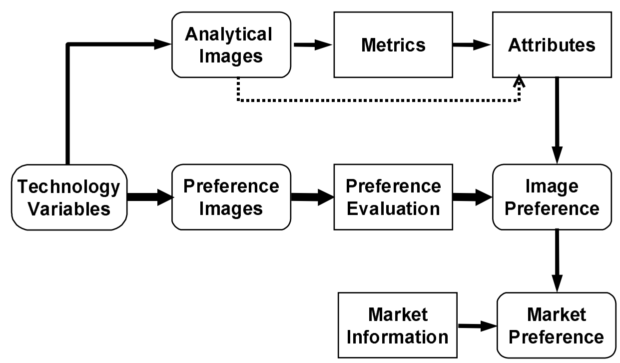

Figure 5: Intended relationship between Technology Variables and Market Preference, in terms of Metrics, DAC Attributes, and Preference.

Preference surveys conducted on customer-like images are used to measure customer image preferences for each of a set of images, and these are combined to measure the overall market preference. Market information should be taken into account in determining the set of preference images to use, and also in the weighting of the individual images to get the market preference.

The measured image preference and measured market preference can be used to develop statistical correlations that will enable prediction of image preference and market preference from the DAC attributes. In principle, once these correlations have been developed from preference data, image preference should be predictable from instrumentally measured image quality metrics, without dependence on human observers. Much work still needs to be done before this scenario can be realized, but such a system could be a powerful tool.

The concept of modeling the overall preference of an image in terms of its image quality attributes is not new. For example, Engledrum[5] has shown that the judged image quality of a set of prints was well predicted by a statistical model incorporating three attributes.

Application of Image Quality Attributes

There are two important applications for image quality attributes. First, they provide a concise and useful way of describing the overall image quality of a given printing system. Secondly, as discussed above, they form an excellent basis for building image preference models.

The DAC system has been invaluable at Xerox for comparing overall image quality of different print engines and image processors. However, at present there are no satisfactory means of communicating image quality with customers or between companies. A system of standard image quality attributes like the DAC system, generally agreed upon by the imaging community and accompanied by specifications for determining all the attributes, could be of significant value to the entire industry.

Summary

Image quality metrics are well defined procedures for quantitatively measuring specific image quality features. Image preference is a measure of how well customers would like a particular image. Both metrics and preference have certain inherent drawbacks as representations of overall image quality of a printer system. A set of image quality attributes, which are high-level image quality descriptors, provides a means of describing the overall image quality of printing systems and is a basis for building image preference models.

References

1. T. Cornsweet, "Visual Perception," Academic Press (1970).
2. B. Wandell, "Foundations of Vision," Sinauer Associates (1995).
3. D. R. Rasmussen, P. A. Crean, F. Nakaya, M. Sato and E. N. Dalal, "Image Quality Metrics: Applications and Requirements," *Proc. PIC Conference*, Portland (1998).
4. R. Balasubramanian and E. Dalal, "A method for quantifying the color gamut of an output device", *Proc. SPIE* **3018**, p. 110-116 (1997).
5. P. Engeldrum, *J. Imaging Sci. and Tech.*, **39**, p. 312-318 (1995).