# Image Quality Modeling: Where Are We?

***Peter G. Engeldrum***
***Imcotek***
***Winchester, MA 01890***
***Pge@imcotek.com***

## Abstract

This review is focused on image quality model building, particularly in the context of the Image Quality Circle. There are two fundamentally different ways to modeling image quality; the impairment approach and the quality approach. Impairment looks at decreases in image quality from some reference or ideal. The quality approach attempts to model the judgment of image quality directly, independent of the reference. The more successful models are called perceptual models, and have perceptual attributes, the ness, as the dependent variables. Generalized weighted mean, or Minkowski metrics, are the most successful mathematical forms of image quality models. Several issues impeding implementation of image quality models remain; appropriate psychometric scaling of quality and nesses, and identification of the nesses, particularly for image coding, compression and processing applications. The Universal Image Quality Model is not on the horizon.

## Introduction

The idea of "the quality of the image" started with the invention of the earliest optical instruments, the optical telescope and microscope (1600-1620). (Galileo was a key figure in both these inventions.) This concept appears again in the early days of photography, 1860-1930, during the development of television, 1935-1955, and continues with digital imaging to the present day.

One might assume that with over four centuries of experience with the concept of image quality that we would be close to a complete understanding of the problem. One reason why we are still far from a complete understanding of image quality, and particularly Image Quality Models, is because we lack a structure or a framework. To address this deficiency a concept called the Image Quality Circle (IQC) was proposed in 1989 at the IS&T Annual Meeting[1].

## The Image Quality Circle

The Image Quality Circle, (IQC) which is shown in Figure 1, is briefly described.

The goal of an imaging system designer is to relate the Technology Variables of the imaging system or technology to the Customer Quality Preference. Figure 1 shows this fundamental objective via the arrow. The link between customer quality preference and the imaging system and materials technology variables is typically determined by selecting a variable, printing images, and then asking customers to judge the quality of the printed image. This clearly works, but it is inefficient over time because a new data collection effort is required every time a parameter is changed. The IQC breaks the relationship between Technology Variables and Customer Perceptions down into a series of definable and measurable steps. The four elements of the IQC approach are depicted in Figure 1 and are described in counter-clockwise order around the Circle.

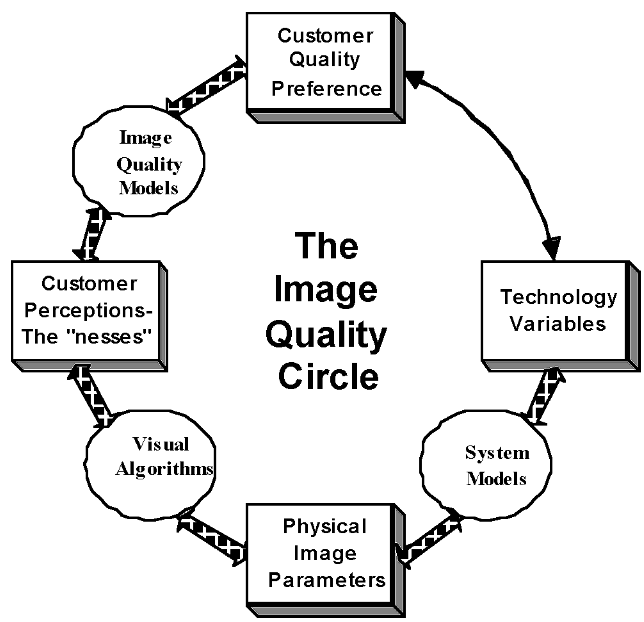


*Figure 1. Image quality circle*

**Customer Quality Preference.**

Customer image quality preference is the overall image quality rating as judged by customers. This is an interval scale of overall image quality that can be defined numerically, say 0 to 100, or adjectivally, such as "bad", "good", or "excellent".

**Customer Perceptions.**

The major customer perceptual attributes of image quality are such dimensions as darkness, sharpness, and graininess. These are called the "nesses" to emphasize the perceptual, as opposed to the physical nature of these attributes.

**Physical image Parameters.**

Physical image parameters are the quantitative functions and parameters we normally ascribe to image quality, such as modulation transfer, Wiener spectra, density, and color.

**Technology Variables.**

Technology variables are the elements or parameters that the imaging system designer or imaging system manufacturer manipulates to change the image quality. Such variables include dots per inch (resolution), toner size, and paper parameters, to name just a few. For any given imaging technology, the list of technology variables is quite extensive.

The four elements of the IQC are linked to one another via models, or algorithms, which are depicted as ovals in Figure 1.

**Image Quality Models.**

Image quality models are empirical (statistical) models that relate the customer perceptions-such as darkness, sharpness, graininess and raggedness, the nesses-to Customer Quality Preferences of image quality. The model describes in mathematical terms the tradeoff that the customer makes when judging image quality. Image Quality Models, as defined here is the topic of this paper.

**Visual Algorithms.**

The algorithm is the recipe that is used to compute a value of a ness from a physical image parameter. An example might be sharpness from the physical measurement of the gradient of a printed edge.

**System Models.**

System models are analytical models that predict the physical image parameters from the technology variables. One example might be the model for the amplitude spectrum of a line boundary (the physical image parameter from which raggedness is calculated) for a dot-matrix printer, developed by knowing dot diameter and dot spacing parameters.

**Terms**

In this discussion, we use the term *image* to mean a colorant arranged in a manner to convey "information." Colorant is used in its most general sense. It can be ink, plastic (toner), wax, dye, silver, phosphors, etc. The image is used to visually communicate information which can be in the form of text, graphs, graphics, images, and art. The idea of an image is very broad and need not be a "hard copy" on a physical substrate. It can be a "soft copy" image on some form of electronic display, or any other appropriate medium.

*Quality* is the integrated set of perceptions of the overall degree of excellence of the image. The set of perceptions can be defined or undefined. For example, in medical imaging, the quality relates to the diagnostic capability of the image, and there are specific protocols for making this evaluation. In most other areas of imaging, image quality is typically a "beauty contest".

The formalism that connects the nesses to the image quality judgment is variously called image quality measures, metrics, and models. It will help to clarify these terms.

## Measures, Metrics and Models

The three M's of measures, metrics, and models have a long and confusing history as applied to image quality so some clarifications are in order.

*Image Quality Measure*—is a signed scalar associated with a vector indicating both its magnitude and sense but not its orientation.

*Image Quality Metric*—posses a distance function that satisfies the triangle inequality.

*Image Quality Model*—a fragment of a mathematical or formal theory of (visual) perception that enables a prediction of image quality from the perceptual attributes that comprise the image quality. Image Quality Models have a specific definition for this review, and that is the one defined in the Image Quality Circle. Here we will focus our attention on the details of connecting the "nesses" to the overall quality judgement.

To compound the three-M confusion, two additional terms have been commonly used in the description of image quality measures: *objective image quality* and *subjective image quality*. Objective image quality measures are usually physical measures, via an instrument, of some image characteristic that is related to the overall quality. In the IQC context, these are termed Physical Image Parameters.

Subjective image quality uses a human being as the measuring instrument. Subjective image quality is often viewed as inferior measurement method compared to objective image quality methods. From a precision or accuracy perspective this may or may not be true, but it misses the fundamental point that humans are the "customer" for images and, by some definition at least, their view of image quality is the correct one.

## The Two Views of Image Quality

There are two dissimilar views of image quality. These two views are predicated on the existence, or convenient availability, of some "original image" or a clearly defined physical limit. One can argue that these two views are fundamentally different. In one instance, the quality of the original image is already built in, and the emphasis is on what degrades or impairs the image quality. The alternative view emphasizes image quality directly, not the degradation or impairment of quality.

For example, the view of, television, image coding compression, and processing, is that there is an "ideal", or a reference image. And in optics there is a physical limit to the quality of an image, diffraction.

The alternative view comes from photography, and digital imaging systems. In this view, not always formally stated, there is no ideal or reference image, except perhaps that in the observers mind. One consequence is that image quality is not bounded, it is open-ended.

These two differing historical views have resulted in two differing approaches to measuring image quality. Television, a standardized system with physical bounds, possesses this concept of an ideal image. It is simply the image quality that is delivered by the system when it is "up to spec". The imaging system performing to specification has a physically describable quality boundary, with deviations measured in terms of impairments. Today we find the identical concept in image quality as applied to image coding, compression, and processing.

A reference, or standard, in optics had its beginnings when astronomer George B. Airy gave the formulation for the diffraction pattern of a clear circular aperture in 1834; the "Airy disk". This physically imposed limit became the measure of the ultimate image quality. Optical Image Quality deviations were first proposed in 1902 when K. Strehl defined the first "image quality measure", what is now known as the Strehl intensity ratio(2). This led to other optical image quality measures such as, "image fidelity", a mean-squared-error criterion between the reference and the reproduced image, the "relative structural content", and the "correlation quality"(2). Most optical and digital imaging and coding quality related measures in use today are related to these measures(3).

Photography, not having a known physical image reference, viewed the problem of image quality in terms of selection of the technology variables to achieve some image quality level. Early photographic technologists recognized the limiting factor of the camera lens, but the image quality capability of the lens was not of prime concern in the development of photographic materials. In fact, no satisfactory way of including the contribution of the lens' imaging characteristics to image quality was available until after World War II. The application of linear systems theory to photography enabled the analysis and integration of lenses and photographic emulsions to optimize system image quality.

## Image Quality Model Theory

### Psychological Underpinnings

According to the framework of the Image Quality Circle, the purpose of an Image Quality Model is to predict the image quality judgment from the value of the nesses in the image. At a very basic level, this is no different from what one does every day. We take in "information" from the world around us via our senses and make decisions based on that information. This is an active research topic in psychology and psychophysics and is termed information integration(4,5) or multidimensional psychophysics(6).

The multidimensional aspect of image quality is, in our context, the nesses or dimensions that drive the image quality judgement. In the psychology literature the Image Quality Model is termed, variously, composition rule(6), combination rule(5), integration model(4) for multidimensional stimuli. Some authors have identified two different types of combination rules(5). They distinguish between the stimulus rule and the perceptual rule. The Image Quality Model, which is a combination of nesses, constitutes a perceptual model (rule). More traditional model building attempts using Physical Image Parameters to predict image quality are categorized as a stimulus model (rule).

Attributes of image quality, the nesses, are either integral or separable. Integral dimensions, or nesses, occur when two dimensions together are perceived as new dimension or percept(5). Separable dimensions are perceived the same when in combination with other dimensions. Image quality, per se, is probably an integral dimension, like color. However, the nesses used in successful Image Quality Models are more than likely separable.

There is little in the psychological literature to choose among for providing a theoretical substrate from which to formalize an approach to Image Quality Models. The most useful approaches have been developed by the imaging community itself.

### Getting the Image Quality Numbers - Psychometric Scaling

The whole of Image Quality Models revolves around numbers representing human judgements of image quality and the nesses. Collecting appropriate human judgements falls into the province of psychometric ("mind measuring") scaling(7).

Although psychometric scaling has a long history in the photographic industry, it is not widely practiced in other areas of imaging. This is unfortunate because appropriate application of psychometric scaling principles is key to precise measurement of the nesses and image quality; the building blocks of Image Quality Models. Some scaling issues will be address later.

### IQ Model Formalism

Many Image Quality Models for imaging has been developed using linear and polynomial regression models on linear or logarithmically transformed variables. The independent variables used in these models have often been the Physical Image Parameters. These have been reported to be practically useful to various degrees(8,9,10,11,12). These models are typically not impairment models.

In the television, digital image compression, and encoding arena a widely used method is the impairment method proposed and developed by Allnatt(13) and colleagues. This is an impairment model and is embodied in an ITU Recommendation BT.500(14). This model starts with the reference image and rates the factors that impair the image. These impairments are additive (subtractive) in their effect on overall picture quality(15).

A variation on the impairment theme, developed by Miyahara, Kotani and colleagues(16,17), uses distortion factors and principle component and multiple regression analysis to construct a Picture Quality Scale. The distortion factors, that include characteristics of the Human Visual System, are developed from the difference image representing before and after encoding.

By far the most successful Image Quality Model formalism in photography, printing, and CRT display is the Minkowski and related metrics. The use of Minkowski metrics has its roots in multidimensional scaling(18) where it is used as a distance measure. The first successful application of the Minkowski metrics to Image Quality Model building, as far as I can determine, was by Bartleson in 1982(19). There were two keys to this success. The first was a break with tradition by using nesses as the independent dimensions (variables) of quality. The second was the choice of separable nesses (dimensions); sharpness and noislessness (10-graininess). In Bartleson's Image Quality Model the Minkowski metric integrated both the combination rule and metric properties into a perceptual model (rule).

A distance interpretation of the Minkowski metric formalism is not the only possibility. It can be generalized and cast as a generalized weighted mean hypothesis (GWMH)[20], suggesting that observers take some form of average when evaluating image quality.

The application of this mathematical formalism has been successful in both the image impairment and quality approachs. Some nesses incorporated into successful image quality models include graininess and sharpness[19,21], defect-ness, sharpness, and color accuracy-ness[20], bluriness and raster-ripple in image coding impairment[22,23,24].

There are at least two reasons for the success of the Minkowski and the GWMH formalism. Two nesses of fundamental importance in photographic, as well as other imaging technologies, is graininess (uniformityness) and sharpness. It appears that these two nesses or dimensions are separable and are represented in "ness space" as two orthogonal axes. Separability of nesses increased the prospect of "finding" a useful model with this formalism. This was serendipitous.

The second reason is the flexibility of the Minkowski and GWMH formalism. The GWMH form tends to mimic the tendency of observers to "peak pick"; i.e. they focus on the worse ness to make their IQ judgment. The magnitude of the exponent captures this.

In the psychology literature a lot of effort has been focused on two different Minkowski metrics, the "city block metric", or linear model, and the Euclidean metric. They only difference between these two, from the mathematical formalism view, is the value of the exponent; one for the city block and two for the Euclidean. In fact one can make an argument that if the nesses are treated as random variables, then small variation in these variables will yield a linear model[20]. These mathematical forms are quite flexible, providing the nesses (dimensions) are separable or independent, the concept of a distance is appropriate, and, observers use the same combination rule.

## Image Quality Model Issues

### Psychometric Scaling

Scaling is key to determining the numerical values of the independent and dependent variables for Image Quality Model building. There are three main issues revolving around scaling. Of major importance is the generation of psychometric scales that have at least an interval, or distance, property. For instance, categorizing responses into N categories and assigning numbers 1 to N to the categories does not guarantee an interval scale.

The second issue is the question posed to the observers. It is common to ask observers for preference of images when the real question is "the quality of the image samples". The unstated assumption is that preference equals quality, which may not be true.

Sample set selection is the third issue. Asking observers to judge quality using a sample set having only one ness varying will usually result in the conclusion that one ness equals quality. And this is just the tip of this iceberg.

### Nesses

The more successful Image Quality Models are ness-based (perceptual) models using the Minkowski or GWMH formalism. (This is the primary reason for the present form of the Image Quality Circle.) In general, it is difficult to identify nesses apriori. Exceptions are the well-known nesses originating from photography: uniformityness (graininess), sharpness, lightness (tone) reproduction-ness, and hue-chroma reproduction-ness. Nesses associated with image coding, compression, and processing are not well understood and more work is needed.

Nesses need to vary for observers to judge quality. Attributes that are visible and detract from the overall image quality, but do not vary, are irrelevant attributes, relative to the IQ judgment.

To date the available data suggest that Image Quality Models built on nesses, that is perceptual models, yield lower variance in the prediction of IQ than models based on Physical Image Parameters. Of course, if the relevant nesses for model building are not known then the only alternative may be the stimulus or PIP form of the IQ models.

### The Universal Image Quality Model

The Universal Image Quality Model has been the Holy Grail of image quality modeling. At present, it is unlikely that the UIQM will ever be achieved. In practical applications of Image Quality Models, the set of relevant nesses is limited to those of interest, and those exhibited in the sample set. This ness set is by no means universal, though a few elements may be common to several imaging technologies. Sharpness is one ness that appears across many facets of imaging.

One could consider a UIQM that has N nesses that encompass some large number of imaging technologies. If N is greater than about five, it is unlikely that observers can attend, simultaneously, to all the N dimensions. Irrelevant nesses are neither part of the judgement process nor the model. This admits the possibility of an array of Image Quality Models that are composed of different sets of nesses.

## Summary

A review of Image Quality Model building reveals that this field is still in its infancy. There is much work to be done!

## References


1. Engeldrum, Peter G., A new approach to image quality, *Presented at the IS&T 42nd Annual Meeting, May 14, 1989, Paper Summaries, pg. 461.* b) Engeldrum, P. G., Measuring customer perception of print quality, *TAPPI Jour.* **73**:161(1990),
2. Linfoot, E. H., Fourier methods in optical image evaluation,

The Focal Press, London, 1964.

3. Eskicioglu, A. M., & P. S. Fisher, Image quality measures and their performance, *IEEE Trans. Comm*. **43**:2959(1995)
4. Massaro, D. W., & D. Friedman, Models of integration given multiple sources of information, *Psychological Review*, **97**:225(1990)
5. Anderson, N., Information integration theory: A brief survey, in Measurement, Psychophysics, and neural Information Processing Volume II, Eds. Krantz, D. H, Atkinson, R. C, Luce, R. D., and Suppes, P, W. H. Freeman & Co., San Francisco, 1974, ISBN 0-7167-0849.3
6. Baird, J. W, Sensation and Judgment: Complementarity Theory of Psychophysics, Lawrence Erlbaum Associates, Mahwah, N. J., 1997 ISBN 0-8058-1830-8
7. Engeldrum, P. G., Psychometric Scaling: A Toolkit for Imaging Systems Engineers, Imcotek Press, 1999.
8. Higgins, G. C., Image quality criteria, *Jour. Appl. Photo. Eng*. **3**:53(1977)
9. Sawyer, J. F., Effect of graininess and sharpness on perceived print quality, *Proceedings of Photographic Image Quality - A symposium held at the University of Oxford,* Sept 8-11, 1980 by the Royal Photographic Society
10. Hunt, B. R., Power-law stimulus-response models for measures of image quality in nonperformance environments, *IEEE Trans. Syst. Man & Cyber*. SMC-9:781(1978)
11. Inagaki, T., Takashi, M., Sasahara, S., Matsuzaki, T., & Goth, T., Color image quality prediction models for color hard copy, *IS&T/SPIE Symposium on Electronic Imaging,* 1994
12. Leachtenauer, J. C., Malila, W., Irvine, J., Colburn, L., & Salvaggio, N., General image-quality equation: GIQE, *Appl. Opt.* **36**:8322(1997)
13. Allnatt, J., Transmitted-picture Assessment, John Wiley & Sons, Chichester, England, 1983 ISBN 0-471-90113-X
14. ITU-R BT.500-7, *Methodology for the subjective assessment of the quality of television pictures*, (1995), http://www.itu.int
15. Allnatt, J., Subjective quality of television pictures with multiple impairments, *Elect. Letters*, **1**:187(1965)
16. Miyahara, M., Quality assessments for visual service, *IEEE Comm. Mag.* **26**:51(1988)
17. Miyahara, M., Kotain, K., and Algazi, V. R., Objective picture quality scale (PQS) for image coding, *IEEE Trans. Comm.* **46**:1215(1998)
18. Shepard, R. N., Attention and the metric structure of the stimulus space, *Jour. Math. Psych*. **1**:54(1964)
19. Bartleson, J., The combined influence of sharpness and graininess on the quality of colour prints, *Jour. Photo. Sci.*, **30**:33(1982)
20. Engeldrum, P. G., A framework for image quality models, *Jour. Imag. Sci. Tech.* **39**:312(1995)
21. Bowman, J., Bullitt, J., Cottrell, F. R., & Hultgren, B., Color print quality II: The combined influence of sharpness & grain, *Summaries, SPSE 42nd Annual Conference*, May 14-19, 1989, pp 466-467, SPSE, Springfield, VA 22151
22. De Ridder, H., Subjective evaluation of scale-space image coding, *SPIE* Volume **1453** *Human Vision, Visual Processing, and Digital Display II* (1991), pg. 31
23. De Ridder, H., Minkowski-metrics as a combination rule for digital-image-coding impairments, *SPIE* Volume **1666** *Human Vision, Visual Processing, and Digital Display III* (1992), pg. 16
24. Nijenhuis, M. R. M, & F. J. J. Blommaert, Perceptual error measure for sampled and interpolated images, *Jour. Imag. Sci. & Tech.* **41**:249(1997)
25. Nijenhuis, M. R. M, & F. J. J. Blommaert, Perceptual error measure and its application to sampled and interpolated single-edge images, *Jour. Opt. Soc. Amer. A* **14**: 2111 (1997)