Preference in Image Quality Modeling

Karin Töpfer, Brian Keelan, Scott O'Dell, and Robert Cookingham
Eastman Kodak Company
Rochester, New York, USA

Abstract

In this paper, we introduce a framework for the inclusion of preference in image quality modeling. The dependence of quality on preferential attributes is expressed as a convolution of two functions, the preference distribution and the quality loss function. The preference distribution is a probability density function characterizing the preferred level of an attribute for a collection of observers and scenes. The quality loss function describes the decrease in quality as the attribute deviates from its observer- and scene-specific preference position. These two functions may be used to predict the value of market segmentation and customization.

Introduction

A variety of image attributes may contribute to the perception of overall quality. We can classify most of these attributes as being either artifactual or preferential in nature. Artifactual attributes are those that generally degrade image quality when they are detectable in an image; examples include graininess, redeye, and aliasing. In contrast, preferential attributes, such as those related to color and tone reproduction (contrast, color saturation, etc.), are always visible in an image, but have an optimal (preferred) position, which may vary as a function of observer and scene.

While it is generally accepted that the changes in perceived image quality resulting from modifications of one or more artifactual attributes can be described in terms of an objective metric, (i.e., a mathematical function of one or more physical quantities that can be measured using a suitable test target), researchers may doubt whether such rigorous modeling can be done for preferential attributes. Preference is considered a matter of personal taste, aesthetics and experience, all of which are hard to quantify. Nonetheless, because color and tone reproduction are an integral part of image quality it is important to integrate artifactual and preferential attributes into an overall framework of image quality.

In the next section, we summarize a mathematical treatment of preferential attributes, which is described in greater detail in Ref. 1. This theoretical description is followed by practical examples of the quantification of preference.

Preferential Attributes in Image Quality Modeling

JNDs of Preference

As described previously,[2] the concept of a just noticeable difference (JND) is central to our image quality framework. It allows us to describe the effect of all attributes on image quality in similar terms so that they can be rigorously combined into a prediction of overall quality in the presence of multiple attributes.

To understand the concept of JNDs in the context of artifactual and preferential attributes, consider the following two examples of paired comparison experiments. In the first experiment, observers are presented with two images, which only differ in noisiness. The observers are instructed to choose which image they perceive to be of higher image quality. If the noisiness difference between the two samples is such that 75% of the observers "correctly" identify the less noisy sample as having higher quality, we define the stimulus difference to be a 50% JND of image quality. The 50% designation is used because it may be inferred that 50% of the observers actually detected the difference in the artifact (noise) level, and chose the less noisy sample, whereas the remaining 50% of observers did not detect any difference between the samples, and so had to guess. By chance, half the guesses would be correct, so a 75%:25% proportion results.

The second paired comparison experiment is similar to the first except that the samples differ only in contrast, which is a preferential attribute, having an optimal value for a given scene and observer. Images of low contrast may appear flat and lifeless, whereas high-contrast images may seem harsh and may lack shadow and/or highlight detail. Despite the fact that the difference in contrast between the samples might be evident to every single observer, not all observers would identify the same sample as being of higher quality, because preference is partly a personal matter. (Similarly, if the same observer evaluated a series of paired images, they would not always prefer the same contrast level, because preference is also scene-dependent.) If 75% of the observers were to select the position of lower contrast as having higher quality, the outcome of the experiment would be the same as that of the first experiment involving the artifact of noise. It is natural to assign an equivalent stimulus difference (i.e., number of JNDs) to samples producing the same paired comparison outcome; hence, the lower contrast sample could be identified as being one 50% JND higher in quality. Even though the word “noticeable” in the term “just noticeable difference” is not strictly appropriate in the context of preference, this terminology is retained for convenience.

In summary, a JND of preference is defined as a stimulus difference producing the same outcome in a paired comparison experiment, as would a JND of an artifactual attribute.

Preference Distributions and Quality Loss Functions

The quality loss function for an artifactual attribute is the relationship between an objective metric value correlated with that artifact and the quality change associated with the presence of that artifact. The quality change is usually referenced to the state in which the artifact is not detectable, and so does not influence quality; consequently, quality changes are normally negative quantities, with zero JNDs of quality change corresponding to a subthreshold level of the artifact. We have fit the quality loss functions of many artifacts using the integrated hyperbolic increment function (IHIF), a simplified form of which is given in Eq. 1.

$$\Delta Q(\Omega) = \frac{R_r}{\Delta\Omega_\infty^2} \cdot \ln\left(1 + \frac{\Delta\Omega_\infty \cdot (\Omega - \Omega_r)}{R_r}\right) - \frac{\Omega - \Omega_r}{\Delta\Omega_\infty} \qquad (1)$$

Here $\Delta Q(\Omega)$ is the quality change (negative) at an objective metric value of Ω, Ω_r is the objective metric value at the reference position (threshold), R_r is the radius of curvature at the reference position, and $\Delta\Omega_\infty$ is the asymptotic objective metric change corresponding to one JND well above threshold.[2] This form of the IHIF applies when $\Omega > \Omega_r$; elsewhere $\Delta Q = 0$.

The quality loss at a given value of an objective metric is a function of both the scene and observer. For example, noise in an image with a large uniform area (such as blue sky) may be more visible and more detrimental to quality than in other types of scenes. A scene for which the quality depends more strongly on an attribute is said to be more susceptible to that attribute. Similarly, different observers may be more or less sensitive to particular attributes. Variations in scene susceptibility and observer sensitivity may be characterized by separately fitting the quality loss data from different subsets of scenes and observers.

The impact of preferential attributes may also be partially described in terms of quality loss functions using Eq. (1), but the reference position is interpreted as the scene- and observer-specific optimum, rather than a threshold of detection. To fully characterize preference, a second function, the preference distribution, is needed. The preference distribution is a probability density function quantifying the relative frequency with which different values of an objective metric are preferred. For example, the preference distribution of contrast for a set of observers and scenes would probably show a peak at some intermediate value of contrast, which provided a good compromise position, and would tail off at lower and higher values of contrast. The preference distribution quantifies how likely it is that a given position be preferred, whereas the quality loss function quantifies how quality falls off away from the optimum position. Formally, the preference distribution of an artifactual attribute may be considered a delta function centered at some subthreshold position, because artifacts are generally preferred to be undetectable.

If the preference distribution and the quality loss function are uncorrelated, then the rapidity of quality loss away from an optimum does not depend upon the position of the optimum. In this case, the mean quality change for a set of observers and scenes at a particular objective metric value is given by the convolution of the preference distribution, denoted $h_p(\Omega)$, and the quality loss function $\Delta Q(\Omega)$.

$$\overline{\Delta Q}(\Omega) = \int_{-\infty}^{+\infty} h_p(\Omega') \cdot \Delta Q(\Omega' - \Omega) \cdot d\Omega' \qquad (2)$$

We have generally found the assumption of a lack of correlation between the preference distribution and quality loss function to be justified, but a minor counter-example is provided by foliage reproduction, where observers preferring more saturated foliage were slightly less sensitive to deviations from their preferred positions.

The implications of Eq. (2) may be easily understood by way of a simple analytical example. If the quality loss function is assumed to be parabolic, rather than being described by the more complicated Eq. (1), and the preference distribution is assumed to be Gaussian (normal), it may be shown that the convolution of Eq. (2) yields the simple result

$$\overline{\Delta Q}(\Omega) = \Delta Q_1 \cdot \left((\Omega - \overline{\Omega}_p)^2 + \sigma_p^2\right) \qquad (3)$$

where σ_p is the dispersion of the Gaussian, Ω_p is the best compromise objective metric value, and ΔQ_1 is the (negative) curvature of the quadratic quality loss function, which is equal to the quality change when the objective metric value is one unit different from the optimum position.

From Eq. (3) it is seen that the mean quality of a given objective metric position is the sum of two terms, the first representing the quality loss arising from the displacement of the position from that of the optimal compromise, and the second representing the quality loss arising because even the best compromise position differs from the individual scene- and observer-dependent optima. The first term can be minimized by careful empirical optimization, leading to proper identification of the best compromise position. The second term can be reduced through customization, i.e., image processing that varies based upon measurable properties of the image and/or available information regarding the preferences of the observer (customer). Alternatively, market segmentation strategies can improve mean quality by providing multiple positions from which an informed choice may be made. The benefit of such an approach can be predicted from Eq. (2) by breaking the convolution down into separate integrals for each position provided.

Example of a Preference Analysis

The Color Balance Experiment

Ten scenes encoded in a balanced RGB scene color representation[3] were employed in this experiment. Before rendering, RGB shifts were applied to the scenes such that the rendition of the 20% gray point represented a central composite design of CIE 1976 a*, b* shifts. Maximum a*and b* shifts of 8 units were selected for all on-axis design points, whereas shifts of 4 units in the a* and b* directions were employed for the diagonal points. The center point, which received no additional shifts, was included, providing nine levels in the experiment. The images were evaluated for overall quality by 22 observers of normal visual acuity and color vision. The softcopy quality ruler workstation employed in the evaluations has been described previously.[4]

Prediction of the Mean JNDs of Quality

Before preference distributions and quality loss functions can be obtained, it is necessary to develop a robust objective metric capable of predicting the data pooled over all judges and scenes. Because of the extensive averaging, the pooled data set has a high signal-to-noise ratio, facilitating the definition of an appropriate objective metric.

A weighted difference of the rendered a^*, b^* values from the average preferred balance position of the 20% gray level, denoted by a_0, b_0, is an intuitively plausible candidate for the objective metric, Ω:

$$\Omega = \sqrt{w_a \cdot (a^* - a_0)^2 + (1 - w_a) \cdot (b^* - b_0)^2} \qquad (4)$$

The weighting factor w_a takes into consideration that equal color shifts in the a* and b* directions may have a different impact on quality. The objective metric, defined in Eq. (4), is inserted into the IHIF, Eq. (1), to fit the mean JNDs of quality for all nine levels. The parameters a_0, b_0, w_a, Ω_r, R_r, and $\Delta\Omega_\infty$ are co-optimized in a nonlinear regression procedure.

Analysis of the Preference Distribution

Our next goal is to analyze the 276 individual data sets for each observer and scene pair to obtain the preference distribution. The simplest way of determining this distribution would be to search for the maximum rating in each data set and to assign the corresponding a*, b* design position as preference. However, the actual preference may fall between any of the design levels, which are relatively widely spaced. Using nonlinear regression methods to obtain the free parameters in Eqs. (1) and (4) can help us to obtain a more accurate estimate of the preference position.

The analysis can be greatly simplified if we assume that the objective metric for the mean and the individual data sets is only a function of the distance from the preferred balance position. This can be achieved by fixing the weighting factor w_a at the position obtained for the mean data set. This is a reasonable assumption given that this weighting factor may be largely determined by human visual system properties, rather than characteristics of individual observers and/or scenes. In the case of color balance, shifts in the b* direction are generally less detrimental to quality than equal shifts in the a* direction, because yellow–blue color shifts are more representative of natural changes in daylight illumination.

While many of the individual data sets are well described by Eqs. (1) and (4), some sets provide little indication of the observer's preference. This might happen if an observer sees only small quality differences between the color renditions of a particular scene. In this case, several levels might receive the maximum rating for quality. Because these particular situations have little impact on the overall preference distribution, the data sets can either be weighted less or totally excluded from the analysis. In most of our studies less than 10% of the sets fell into this category.

We found that the following criteria worked well in determining the consistency of individual data sets: (1) In some cases, an individual rating in the set was inconsistent with all other ratings. Removing the outlier improved the estimate of the preference position. (2) We calculated the correlation coefficient between the measured JNDs and the predictions from Eqs. (1) and (4). Data sets with correlation coefficients below a certain threshold were excluded. (3) In some situations, the correlation fell above the threshold, but the regressed Ω_r parameter in the IHIF was unusually large. This corresponds to a high threshold in the objective metric below, which no quality degradations are observed, indicating larger uncertainties about the actual preference. Consequently, data sets with higher thresholds were weighted less (using a factor of by $1/(1+\Omega_r)$) in computing the preference distributions.

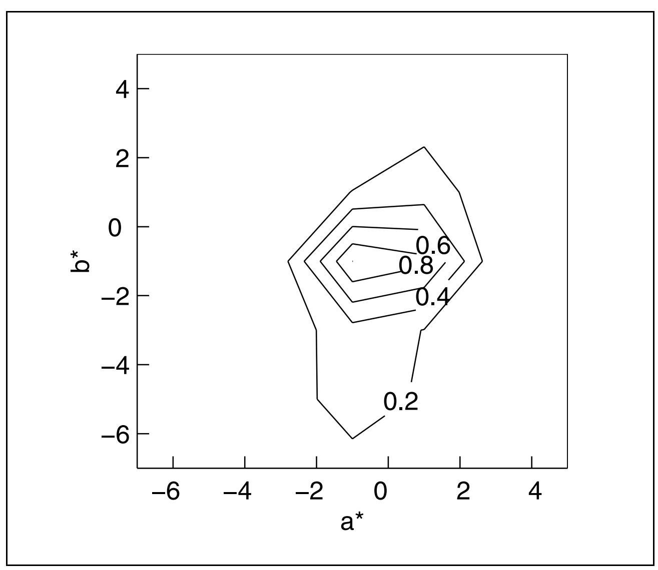

Figure 1. Preference distribution for color balance shown as contours in a CIELAB a vs. b* diagram.*

The preferred a*, b* values a_0, b_0 were recorded for each scene-observer combination together with the regression parameters for the IHIF, Eq. (1). The frequency distribution of preferences was obtained by sorting the preference points into two-dimensional bins with a bin width of 2 a*, b* units and incrementing the count in the appropriate bin by $1/(1+\Omega_r)$ (see item 3 above). The optimum bin size depends on the width of the preference distribution and the number of individual entries available from the experiment. Figure 1 shows the preference distribution obtained from the color balance experiment as a contour plot in the a*, b* plane. The contours represent the normalized weighted frequency distribution. The preference distribution falls off more rapidly in the red-green (a*) direction, and is asymmetric in the blue-yellow direction with a higher probability of preferences in the blue direction.

Analysis of the Quality Loss Function

An initial estimate of the quality loss function can be made by aligning the optima of the individual quality loss functions for all scene-observer combinations. This can be achieved by setting the parameters a_0, b_0 in Eq. (4) to zero. The *a*, b** values in the objective metric, Eq. (4), can now be interpreted as differences from the optimum color balance, which corresponds to zero JNDs of quality degradation according to Eqs. (1) and (4). We can use the stored regression parameters for each data set to calculate the individual quality loss functions. Averaging produces an estimate of the overall quality loss function, which can again be modeled using Eq. (1). The dotted line in Fig. 2 shows the results of this analysis. An average quality loss function, which provides an even better prediction of the mean JNDs, pooled over scene and judge, is produced by minimizing the sum of square errors between measured and predicted mean JNDs, using Eqs. (1) and (2) in a nonlinear regression routine. The preference distribution shown in Fig. 1 corresponds to the quantity h_p in Eq. (2). The dashed line in Fig. 2 shows the quality loss function (ΔQ in Eq. (2), modeled using the IHIF, Eq. (1)) determined by this method.

The quality gain associated with customizing the color balance for each scene and observer, is likewise obtained in the nonlinear regression routine. This quantity corresponds to the term $\Delta Q_1 \cdot \sigma_p^2$ in Eq. (3). According to this study, we would gain 3 JNDs of quality improvement if the color balance was customized for each observer and scene, as opposed to providing an optimum average balance.

A comparison of the quality loss function with the IHIF obtained for the mean JNDs, represented by the solid line in Fig. 2, can help us determine if the average response of the observers is dominated by the quality loss function or by the preference distribution. In most cases, the convolution integral in Eq. (2) will lead to a less rapid fall-off of the mean compared with the quality loss function. However, if the quality loss function is a parabolic, both functions are identical! In the case of color balance, both functions shape the curve for the mean JNDs as a function of the objective metric.

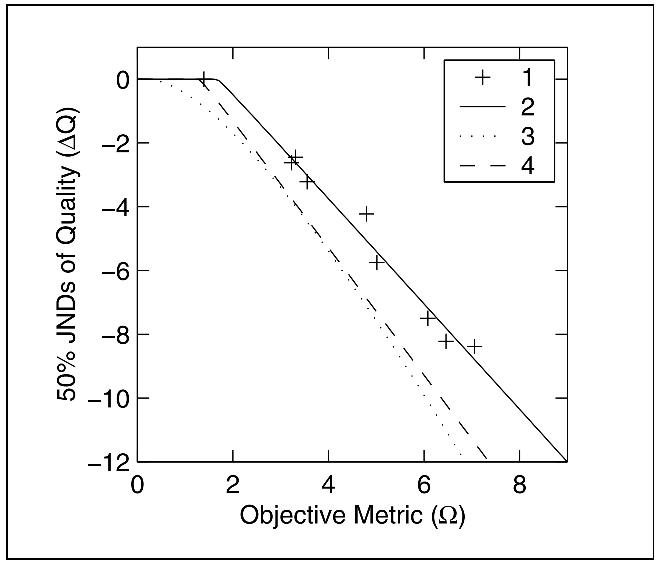

Figure 2. Quality loss functions, computed four ways. (1) Measured data for the mean observer and scene. (2) Regression for mean observer and scene. (3) Average quality loss function from the regression of individual data sets, Eqs. (1) and (4). (4) Quality loss function corresponding to the preference distribution in Fig. 1, Eqs. (1) and (2).

Preference Distributions and Quality Loss for Key Colors

The interplay of preference and quality loss gives us powerful insights into the complexities of our perception of color quality. This becomes evident if the results of the color balance study are compared with results of previous memory color studies.[3]

Figure 3 summarizes the results regarding preference distributions and quality loss functions for neutral colors (color balance) and for the memory colors skin, foliage, and blue sky in a CIELAB plot. The dotted lines refer to the 90% preference contours, corresponding to a 90% fall-off of the preference distribution compared with its maximum value. The solid lines represent the quality loss contours, corresponding to five JNDs of quality degradation compared with the optimum. The positions of the contours correspond to the preferred reproduction of the scene color that most typically represents each color.

Depending on the color of interest, quality loss functions and preference distributions can have distinctly different shapes. This means that equal color shifts in an approximately perceptually uniform color space can have a very different impact on color quality, depending on the color under consideration and the direction of the shift.

It is interesting that the 90% preference contours and –5 JND quality contours have very similar sizes and almost overlap in some cases, e.g., foliage. This suggests that, if expressed in terms of quality fall-off from an overall optimum for each color, the extent of the preference distributions remains almost invariant, regardless of the color under consideration. For example, quality degrades rapidly if the skin tone reproduction moves away from the optimum, and the preference distribution is correspondingly narrow. In contrast, chroma changes in the reproduction of blue sky have a small impact on quality, and the preference distribution is relatively wide in this direction.

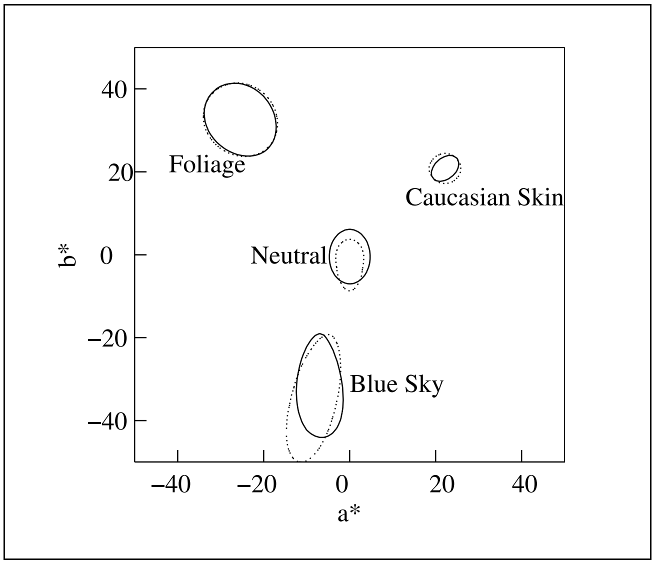

Figure 3. Preference distributions (dotted lines) and quality loss functions (solid lines) shown as contours in a CIELAB a vs b* plot. The preference contour corresponds to 10% of the maximum value of the preference distribution. The quality loss contour indicates a 5 JND quality degradation compared with the optimum. Reprinted from Ref. 4 by courtesy of Marcel Dekker, Inc.*

Conclusion

Artifactual and preferential image quality attributes have been integrated into a framework for the prediction of overall quality based on the concept of just noticeable differences (JNDs). A JND of preference is defined as a stimulus difference producing the same outcome in a paired comparison experiment, as would a JND of an artifactual attribute.

The common element in modeling artifactual and preferential attributes is the quality loss function, Eq. (1), which is the relationship between an objective metric value correlated with the attribute and the associated quality change. In the case of artifacts, the quality change is usually referenced to the state in which the artifact is not detectable, and so does not influence quality. For preferential attributes, the quality loss function describes the decrease in quality as the attribute deviates from its observer- and scene-specific preference position and is also modeled using Eq. (1).

The parameters characterizing the preferred value for a color and tone attribute are part of the objective metric, which is often formulated as a weighted CIELAB color difference from the optimum (e.g., Eq. (4)).

The dependence of quality on preferential attributes may be expressed as a convolution of two functions, the preference distribution and the quality loss function. The preference distribution is a probability density function characterizing the preferred level of an attribute for a collection of observers and scenes.

For several key colors, we demonstrated that their preference distributions and quality loss functions have distinctly different shapes. However, the size of the 90% preference distribution seemed to correspond to an almost constant quality loss of approximately 5 JND from the average optimum position, regardless of the color under consideration and the direction of the shift.

References

1. Brian W. Keelan, *Handbook of Image Quality: Characterization and Prediction*, Marcel Dekker, Inc., New York, Ch. 4 (2002).
2. Brian W. Keelan, “Characterization and Prediction of Image Quality”, *Proc. PICS 2000*, pp. 197–203 (2000).
3. Kevin E. Spaulding, Geoffrey J. Woolfe, and Edward J. Giorgianni, “Reference Input/Output Medium Metric RGB Color Encodings (RIMM/ROMM RGB)”, *Proc. PICS 2000*, pp. 155–163 (2000).
4. Karin Töpfer, Robert E. Cookingham, “Quantitative Aspects of Color Rendering for Memory Colors”, *Proc. PICS 2000*, pp. 94–98 (2000).

Biography

Karin Töpfer received her Masters degree in Physics from Dresden University of Technology in 1983 and a Ph.D. in Photophysics from Dresden University of Technology in 1985. Since 1993, she has worked at Eastman Kodak Company, first in the U.K. and later in Rochester, NY. In recent years, her work has primarily focused on image quality modeling and psychophysics, including color quality. She is a member of the IS&T and a Fellow of the Royal Photographic Society.