Perceptual Error Measure for Sampled and Interpolated Images

M. R. M. Nijenhuis and F. J. J. Blommaert

Institute for Perception Research / IPO, P.O. Box 513, 5600 MB Eindhoven, The Netherlands

Error measures quantify the difference between a reproduced image and the corresponding unprocessed "original" image. Unfortunately, most of the existing error measures such as the mean-square-error (MSE) correlate poorly with the perceived quality of the reproduced images. The reason is that these measures either do not or insufficiently take the properties of the human visual system into account. The distance in a perceptual space spanned by artifacts relevant to the image reproduction techniques is used as a measure of the impairment of the reproduced image relative to the original image. For sampling and interpolation, we show how a two-dimensional perceptual space with the sensorial strengths of periodic structure and blur along the axes can be constructed from the physical parameters. The quantitative perceptual error measure can be used to determine a perceptually optimal combination of sampling and interpolation. The optimization problem is shown to be equivalent to minimizing a cost function known from standard regularization theory. The optimal solution is a compromise between conflicting demands in the perceptual space.

Journal of Imaging Science and Technology 41: 249–258 (1997)

Introduction

An ideal display system produces sampled and interpolated images that are visually indistinguishable from continuous images. However, most display systems are not ideal but produce impaired versions of continuous images. It is therefore important to develop error measures that define the quality of the sampled and interpolated image in comparison with the continuous image. Such measures facilitate the design of display systems that meet a certain quality criterion, and in general are important for the optimization of display parameters.

In this paper, we derive a new perceptual error measure for sampled and interpolated images.[1] We propose that the difference between the continuous version and the sampled and interpolated version of an image be defined by the total perceptual impairment caused by the sampling and interpolation process. The nature as well as the magnitude of the distance measure is, in this case, determined by visual perception. Reference 2 shows that perceptual quality is linearly related to perceptual impairment. Therefore, the perceptual distance measure is also proportional to the loss of perceptual quality due to the sampling and interpolation process.

The perceptual error measure is based on global knowledge about the structure of the perceptual process that leads to quality judgments about images. This new measure models the perceptual processes relevant to quality judgments rather than modeling the functions of (groups of) cells in the visual system, as in Ref. 3. Although the latter approach may work well for low-level vision, it is difficult to model high-level vision (grouping, cognition) in such a way. The global knowledge of the perceptual structure in our metric is mainly based on the accepted notion that the contribution of different impairments to image quality is usually independent and can therefore be described in a multidimensional attribute space.[2,4] Furthermore, we use the conception that the perceptual and cognitive processing of artifacts can be described by a sequence of mappings: from physical parameters to low-level attributes, from low-level attributes to cognitive attributes (impairments), and from impairments to quality. In a companion paper, these steps will be worked out in detail.

In this paper, we derive a perceptual error measure for the impairment of sampled and interpolated images. In related studies, this error measure is validated for single-edged black-and-white images,[1,5] complex black-and-white images, and complex color images.[1,6]

Perceptual Error Measure

In this section, we formulate a quantitative expression for the total perceptual impairment of sampled and interpolated images relative to the originals, i.e., images produced by a hypothetical ideal display. Figure 1 shows the effects of sampling and interpolation for a portrait of a woman, "Wanda." The original image is shown in Fig. 1(a). Figures 1(b), 1(c), and 1(d) are impaired versions of the original in which sampling and interpolation artifacts are visible. Typical artifacts are periodic structure, blur, staircase, and moire. The term "staircase" describes milled edges and lines. By moiré we mean a perceived low-frequency periodic brightness variation caused by the interaction of a periodic sampling structure of relatively high frequency with a periodic brightness variation in the image of approximately the same frequency. Each of the artifacts has its own specific effect on the total perceptual impairment. Periodic structure and blur are the most dominant artifacts in the "Wanda" image. Staircase is only visible in a few local areas of the image. Although artifacts such as moiré are not visible in Fig. 1, they can be perceived in local areas of other images. Figures 1(b), 1(c), and 1(d) also show that the perceptual strength of periodic structure can be reduced by stronger interpolation at the expense of the sharpness of the image.

Figure 2 shows the different computational steps of the perceptual error measure, which determines the perceptual quality Q from the physical parameters Φ_i, specifying both the image and the sampling and interpolation process. The perceptual quality is determined from the total perceptual impairment I. The total perceptual impairment is a distance in a perceptual space. We assume

Original manuscript received April 5, 1996.

Figure 1. Effects of sampling and interpolation: (a) original complex black-and-white image "Wanda," (b) columnar sampling and optical Gaussian interpolation with small spread parameter, (c) ditto with medium spread parameter, and (d) ditto with large spread parameter.

that the artifacts relevant to the sampling and interpolation process are the orthogonal dimensions of the perceptual space. The most dominant artifacts, periodic structure and blur, are denoted by the subscripts p and b, respectively. It is this perceptual space that distinguishes the perceptual error measure front the conventional error measures. The impairments I_i, i.e., the weighted perceptual strengths S_i, of the individual sampling and interpolation artifacts are along the axes of the perceptual space. The sensorial strengths are determined from the physical parameters.

Perceptual Impairment and Perceptual Image Quality. Measurements plotted in Fig. 3 show that the relationship between perceptual impairment and perceptual image quality is linear to a good approximation. This relationship holds for different images, different artifacts, a large quality range, and images with one as well as multiple artifacts.[7] We assume it to hold for all combinations of sampling and interpolation artifacts. The data in Fig. 3 and most of the other figures were obtained from category scaling experiments. In these experiments subjects rated the perceptual attribute of interest by giving a number on a 10-point numerical scale. These data were plotted directly in the graphs. In most cases the data were averaged over subjects and/or images. A detailed description of a scaling experiment can be found in the Appendix. Unless otherwise stated, the length of an error bar in this

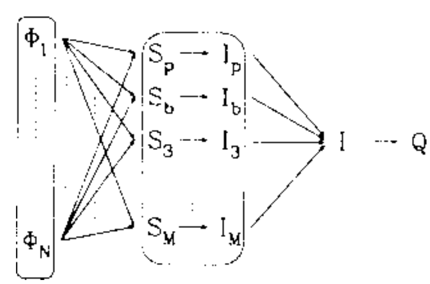

Figure 2. Schematic diagram of the computational steps of the perceptual error measure that relates perceptual image quality (Q) or the total perceptual impairment (I) to the physical parameters (Φ_i), using an intermediate perceptual space with sensorial strengths (S_i) or impairments (I_i) or the individual sampling and interpolation artifacts along the orthogonal axes.

and other figures is twice the standard error of the mean averaged across the data points. The average standard error was plotted to make the plots as simple as possible. The loss of information due to using the average standard error is only minor, since the standard error does not vary systematically across the data points, and deviations from the averaged value are small.

The perceptual impairment I can thus be converted from or into perceptual quality Q by

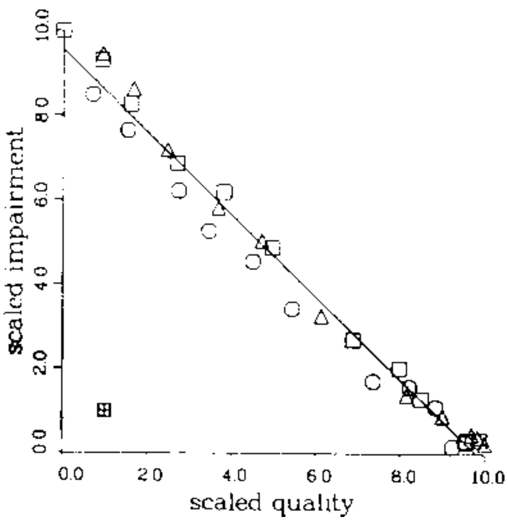

Figure 3. Experimentally determined category scaling data for perceptual impairment versus experimentally determined category scaling data for perceptual image quality for complex black-and-white images impaired by additive white Gaussian noise. The data have been averaged across subjects. Unless otherwise states, the length of an error bar in this and other figures is twice the standard error of the mean averaged over the data points. (□) Image "Wanda"; (○) image "Fruit"; and (Δ) image "Terrasgeel." Similar to the "Wanda" image, the images "Fruit" and "Terrasgeel" are complex images depicting a market stall displaying fruit and a terrace with a yellow parasol, respectively.

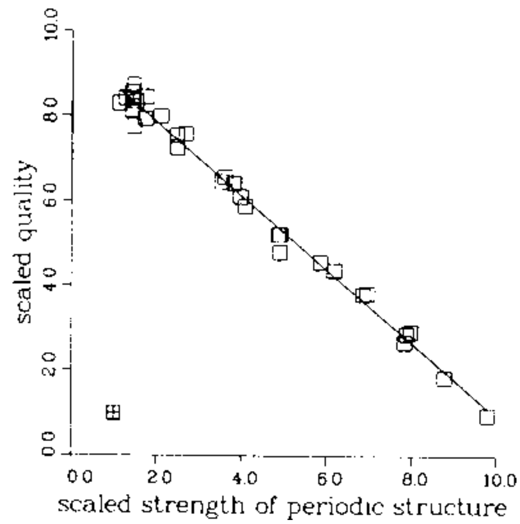

Figure 4. Experimentally determined category scaling data for perceptual image quality versus experimentally determined category scaling data forperceptual strength of periodic structure as measured in the periodic structure experiment described in the Appendix. The data have been averaged across subjects.

$$I = 1 - Q. \tag{1}$$

In this paper, inverse linearly related perceptual attributes such as perceptual impairment and perceptual quality are called complementary attributes.

Note that there is a difference between the perceptual attributes in the model and their measured counterparts. Throughout this paper, we assume that when category scaling experiments are used, the relations between the measured perceptual attributes are identical to the relations between the perceptual attributes except for a linear transformation. For computational convenience, the perceptual attributes used in the perceptual error measure are in the interval [0,1]. The measured perceptual attributes may be in a different interval. Usually, the difference between a perceptual attribute and the corresponding measured perceptual attribute is clear from the context. If this is not the case, we use subscripts to distinguish between the two.

Combination Rule for Perceptual Impairments. Besides judging total perceptual impairment, people can also distinguish between different artifacts that impair the displayed image and judge their perceptual impairments separately. Minkowski metrics[2,8] can be used to combine the M underlying perceptual impairments I_i into the total impairment I:

$$I^{\alpha} = \sum_{i=1}^{M} I_i^{\alpha} = I_p^{\alpha} + I_b^{\alpha} + \sum_{i=3}^{M} I_i^{\alpha}, \tag{2}$$

where the Minkowski exponent α is a parameter to be determined experimentally. Because the artifacts are the dimensions of the perceptual space, the impairment I can alternatively be interpreted as some distance in this space. For $\alpha = 2$, the distance is Euclidian.

Perceptual Strength and Impairment of an Artifact. The perceptual impairment caused by a specific artifact increases as the artifact gets stronger. Measurements described next show that for the artifact's periodic structure (subscript p) and blur (subscript b), the impairments can be related linearly to the perceptual strengths S of the artifacts:

$$I_p = a_p S_p + b_p, \tag{3}$$

$$I_b = a_b S_b + b_b. \tag{4}$$

Obviously, the constants b_p and b_b are zero since there is no impairment if the sensorial strength of an artifact is zero. We assume that this linear relation holds for other artifacts as well:

$$I_i = a_i S_i. \tag{5}$$

The constants a_i carry the relative weights of the artifacts in the total perceptual impairment.

Since perceptual quality on the vertical axis of Fig. 4 can be converted linearly into impairment (Eq. 1), the data from the periodic structure experiment (for details see Appendix), plotted in Fig. 4, indicate a linear relation between impairment and perceptual strength of periodic structure. Similarly, the experimental data from Westerink[9] in Fig. 5 showing a linear relation between quality and sharpness, can be used to illustrate the linear relation between perceptual impairment and the perceptual strength of blur. To this end we have to show that the relation between impairment and quality and that between sharpness and perceived blur are both linear. According to Fig. 3, quality can be converted linearly into impairment. Sharpness on the horizontal axis is the complementary perceptual attribute of perceived blur. Figure 6 indicates that these complementary attributes are linearly related as would be expected from similar data for lightness and darkness and loudness and softness.[10,11]

Physical Parameters and Perceptual Strengths. The strategy introduced by Fechner[12] and recently advocated by

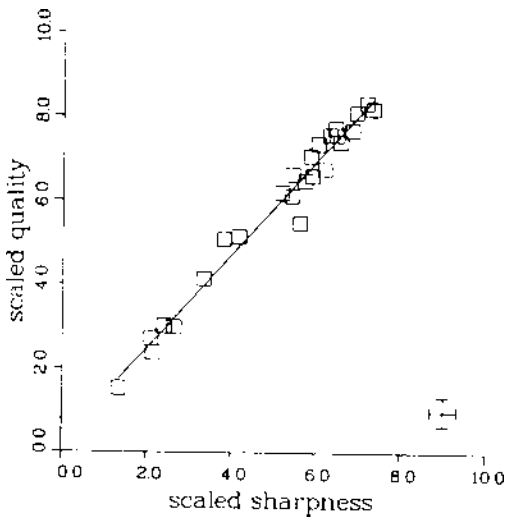

Figure 5. Experimentally determined category scaling data for perceptual image quality versus experimentally determined category scaling data for perceptual sharpness. Each data point is the average of 80 values obtained from 20 subjects each judging four complex images. Adapted from Westerink.[9]

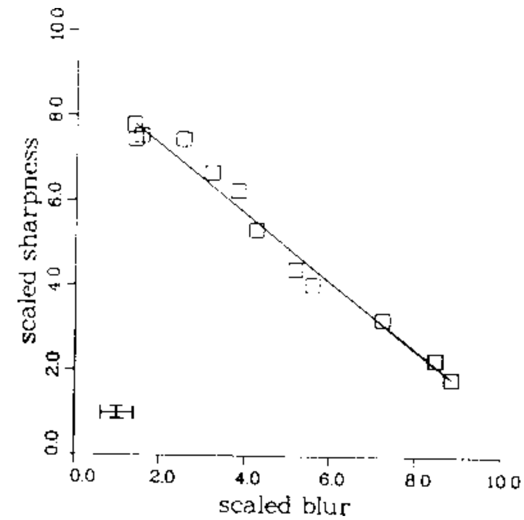

Figure 6. Experimentally determined category ratings for both sharpness and perceived blur for a black-and-white portrait of a woman, "Wanda." The data have been averaged across subjects. The perceived blur data have been published by de Ridder and Majoor.[26] Unpublished sharpness data have been measured by de Ridder under the same experimental conditions. Courtesy de Ridder.

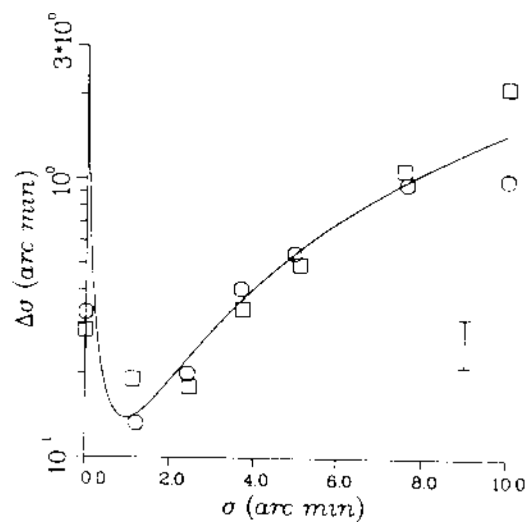

Figure 7. Measured JND for edge blur as a function of the physical Gaussian blur parameter σ for two subjects □ and ❍. Adapted from Watt and Morgan.[16] The solid curve represents calculated JND using a formula obtained from the substitution of Eq. 11 into Eq. 9 with σ_0 = 1.2 arc min and k = 0.03.

Watt[13] and by Wilson[14] has been used to derive quantitative expressions for the sensorial strengths S_b and S_p of the artifact's blur and periodic structure as a function of the physical parameters. The interpolation function is assumed to have a Gaussian impulse response $h(x)$ with spread parameter σ:

$$h(x) = \frac{1}{\sqrt{2\pi}\sigma}\exp\left(\frac{-x^2}{2\sigma^2}\right). \quad (6)$$

The reason for choosing the Gaussian interpolation function is twofold. First, it is frequently used. Second, as can be shown by using the central limit theorem, a combination of various resolution degrading filters produces an overall response that is approximately Gaussian.[15]

We assume that the perceptual strength of blur is a differentiable function of the physical spread parameter σ:

$$S_b = S_b(\sigma). \quad (7)$$

Variations in the perceptual variable S_b are then related to variations in the physical variable σ by

$$\Delta S_b = \frac{dS_b}{d\sigma}\Delta\sigma. \quad (8)$$

According to Fechner, a key property of perceptual attributes is that just noticeable differences (JND) of the perceptual variable are independent of the strength of the attribute: $\Delta S_b = k$. Consequently, the sensorial strength function for blur can be constructed by measuring JND $\Delta\sigma$ as a function of σ and deriving S_b from the equation

$$\frac{dS_b}{d\sigma} = \frac{k}{\Delta\sigma(\sigma)}. \quad (9)$$

Experimental results from Watt and Morgan[16] are adapted in Fig. 7. These experimental data show that, for larger σ values (2.5 ≤ σ ≤ 10 arc min), Δσ *is* proportional to $\sigma^{1.5}$. Therefore, the sensorial strength function for blur can be described by

$$S_b = a\frac{1}{\sqrt{\sigma}} + b. \quad (10)$$

Since this function is not suitable for describing the minimum in the Watt and Morgan[16] data plotted in Fig. 7 for small values of a (0 ≤ σ ≤ 2.5 arc min), we use a slightly modified version of Eq. 10 in which σ is replaced by $\left(\sigma^2 + \sigma_0^2\right)^{\frac{1}{2}}$:

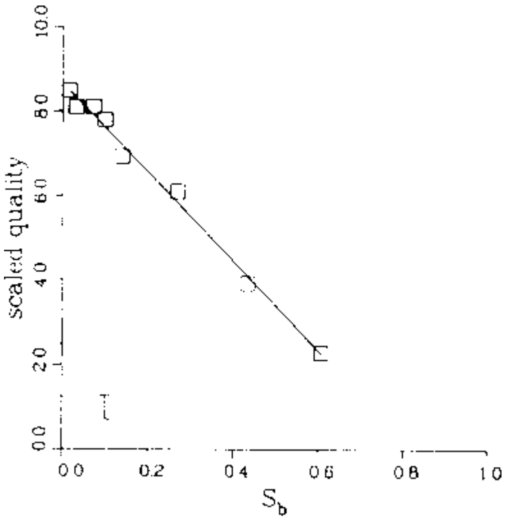

Figure 8. Experimentally determined category scaling data for perceptual quality of Gaussian blurrred complex images versus the calculatd sensorial strength of blur. The experimental data were adapted from Westerink and Roufs.[27] Sensorial strength of blur was calculated from the σ values in the plot of Westerink and Rouf using Eq. 11 with $\sigma_0 = 0.70$ arc min.

$$S_b = 1 - \frac{1}{\left(\left(\frac{\sigma}{\sigma_0}\right)^2 + 1\right)^{0.25}}, \tag{11}$$

For conveniences the constants are chosen such that $0 \le S_b \le 1$. The parameter σ_0 represents the intrinsic blur of the early-visual pathway, which may contain both optical and physiological factors.

The sensorial strength function for blur is consistent with both JND and scaling data for blur. Substitution of the sensorial strength function for blur (S_b) of Eq. 11 into Eq. 9 yields an expression for $\Delta\sigma$ as a function of σ. The dipper-shaped curve in Fig. 7 is a plot of $\Delta\sigma$ as a function of σ for $\sigma_0 = 1.2$ arc min and $k = 0.03$ and fits the JND data of Watt and Morgan[16] well. Figure 8 indicates that the sensorial strength function for blur can be used successfully to predict the scaled perceptual quality of complex images blurred by a filter with a Gaussian impulse response with spread parameter σ.

The sensorial strength function for periodic structure (S_p) is derived similarly. If we assume that the sensorial strength of periodic structure is a differentiable function of the modulation depth m, then S_p can be derived from

$$\frac{dS_p}{dm} = \frac{k}{\Delta m(m)}, \tag{12}$$

provided that we know JND Δm as a function of m. Experimental results from Legge[17] plotted in Fig. 9 and Carlson and Cohen[18] show that, for larger m values, Δm is proportional to m^β. Hence, S_p must be of the form

$$S_p = am^{1-\beta} + b. \tag{13}$$

According to the data from Legge[17] in Fig. 9 and Carlson and Cohen,[18] the value of the exponent β in $\Delta m \propto m^\beta$ increases slightly with the frequency f of the sine grating.

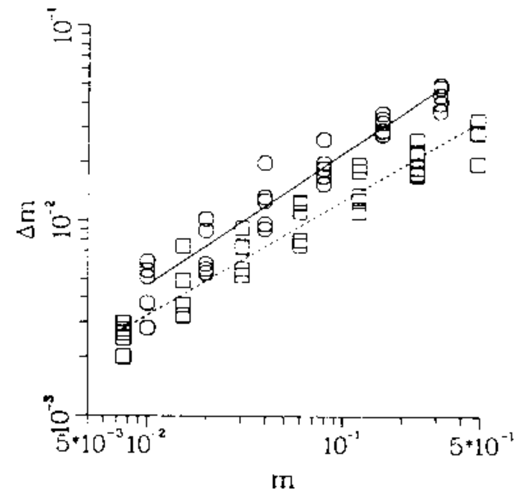

Figure 9. JND of a single subject for modulation depth of sine grating patterns as a function of the modulation depth m with the frequency as a parameter. The □ and ❍ symbols indicate JND data for sine grating frequencies of 2 and 8 cpd, respectively. Each point is the geometric mean of four measurements and the lines are regression lines. The dashed line has a slope of 0.59 and the continuous line has a slope of 0.68. Adapted from Legge.[19]

As a first-order approximation, we assume that the exponent β increases linearly with frequency:

$$\beta = \beta_0\left(1 + \frac{f - f_{\beta_0}}{f_0}\right), \tag{14}$$

where f_0 is a constant and $f_{\beta 0}$ *is* the frequency for which $\beta = \beta_0$. According to Legge, Carlson and Cohen[17,18] and analysis of the experimental data of the periodic structure experiment described in the Appendix, typical values are $\beta_0 = 0.7$, $f_0 = 98.4$ cpd and $f_{\beta 0} = 12.7$ cpd.

For small values of m at threshold, JND Δm are no longer proportional to m^β. Wilson[14] derived a sensorial strength function valid for both threshold and suprathreshold values. In this paper, we use a simplified version:

$$S_p = c\frac{\left[1 + \left(\frac{m}{m_0}\right)^3\right]^{1/3} - 1}{\left(\frac{m}{m_0}\right)^\beta}, \tag{15}$$

where m_0 is a threshold constant that still has to be determined. The constant $c = (1/m_0)^{\beta 0}/\{[1 + (1/m_0)^3]^{1/3} - 1\}$ is chosen such that $0 \le S_p \le 1$ for $f > f_{\beta 0}$. Note that for higher values of m, the perceptual strength is proportional to $m^{1-\beta}$.

The modulation depth is specified by:

$$m = \exp\left[-2\left(\frac{\pi}{d}\right)^2\left(\sigma^2 + \sigma_0^2\right)\right], \tag{16}$$

where $1/d$ is the sample frequency, σ_0 is again the intrinsic blur of the early visual pathway, and σ is the spread parameter of an optical Gaussian interpolation filter. The

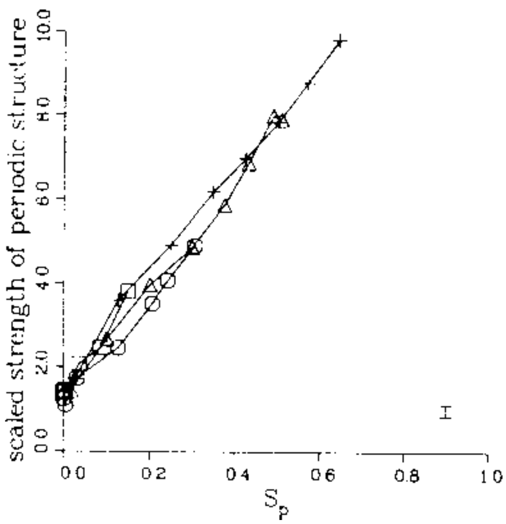

Figure 10. Experimentally determined category scaling data for the sensorial strength of periodic structure versus the calculated sensorial strength of periodic structure with sampling distance as a parameter: (□) d = 1.47 arc min, (○) d = 1.96 arc min, (Δ) d = 2.94 arc min, and (+) d = 3.92 arc min. The experimetnal data were obtained from the periodic structure experiment described in the Appendix. Experimental data were averaged across subjects. The sensorial strength of periodic structure was calculated using Eq. 15. The parameters σ_0 and m_0 were fitted such that all data points are on a straight line: σ_0 = 0.62 arc min, m_0 = 0.018. The remaining parameters were fixed: β_0 = 0.7, f_0 = 98.4 cpd, $f_{\beta 0}$ = 12.7 cpd.

expression is consistent with the high-frequency envelope of experimentally determined contrast sensitivity functions, as in Ref. 19.

Figure 10 shows that the calculated sensorial strength of periodic structure is consistent with experimental scaling data for the perceptual strength of periodic structure.

Interpretation of the Perceptual Error Measure

By substitution of Eq. 5, which relates the perceptual strength to the impairment of an artifact, Eq. 2 for the total impairment can be written as

$$I^{\alpha} = a_p^{\alpha} S_p^{\alpha} + a_b^{\alpha} S_b^{\alpha} + \sum_{i=3}^{M} a_i^{\alpha} S_i^{\alpha}, \tag{17}$$

which is equivalent to

$$I^{\alpha} = a_p^{\alpha} \left[S_p^{\alpha} + \lambda_b S_b^{\alpha} + \sum_{i=3}^{M} \lambda_i S_i^{\alpha} \right], \tag{18}$$

with $\lambda_i = (a_i/a_p)^{\alpha}$. The λ_i parameters tell us something about an observer's weight of the periodic structure artifact relative to the weights of the other sampling and interpolation artifacts. Equation 18 shows that optimization of the sampling and interpolation problem is equivalent to minimizing the function

$$S_p^{\alpha} + \lambda_b S_b^{\alpha} + \sum_{i=3}^{M} \lambda_i S_i^{\alpha}, \tag{19}$$

which may be regarded as a cost function.

Minimization of the cost function permits the specification of optimal physical parameter values that minimize the perceptual impairment. If, for example, the interpolation is assumed to be Gaussian with spread parameter σ, then the sensorial strengths of periodic structure, blur, and other sampling and interpolation artifacts depend on σ. In this case, the optimal σ value can be solved from

$$\frac{\partial}{\partial \sigma} \left[S_p^{\alpha}(\sigma) + \lambda_b S_b^{\alpha}(\sigma) + \sum_{i=3}^{M} \lambda_i S_i^{\alpha}(\sigma) \right] = 0. \tag{20}$$

Minimization of cost functions is a well-known variational regularization solution method for ill-posed problems.[20] The "ill-posedness" often manifests itself as the absence of a unique solution to the problem. Variational principles are used widely in physics, econnomics, and engineering. In physics, for instance, most of the basic laws have a compact formulation in terms of variational principles that require minimization of a suitable functional such as energy or Lagrangian.[20] Variational regularization imposes constraints on the possible solutions of an ill-posed problem to reduce the number of solutions and thus restore its "well-posedness." These constraints often conflict. The solution is then a compromise between several conflicting demands. The regularization parameters λ_i control the relative importance of the constraints.

The sampling and interpolation problem for a uniform image is ill-posed since there exists no unique solution for the interpolation function. Any interpolation function that makes the periodic structure invisible is a solution. The sampling and interpolation problem can be made well-posed by imposing the additional constraint that nonhomogeneous images should also be free from other sampling and interpolation artifacts. The regularization parameters λ_i say something about an observer's weight of periodic structure compared to the observer's weight of other sampling and interpolation artifacts such as blur and staircase. Solutions obtained in this way give the best compromise between the various impairments in the image.

What the best compromise is depends on the regularization parameters or, more precisely, on the relative weights of artifacts compared to the artifact periodic structure. The relative weights are thought to depend on observer properties as well as on image content. Subjective factors such as individual preference influence the relative weights. Image content determines the prominence of an artifact since it influences whether or not the artifact occurs and how often it occurs. For example, only a few images will contain moiré, and a few more will contain the staircase effect. Almost all images will contain blur, but the periodic structure artifact caused by sampling will occur in all images. This is the compelling reason why the cost function should be written as Eq. 19.

The usefulness of the approach depends to some extent on the generalizability of the regularization parameters λ_i over different observers and different images or, more generally, on the generalizability of the optimal sampling and interpolation solution over different observers and different images. In related studies,[1,5,6] we will therefore investigate the behavior of the regularization parameters for: (1) different observers; (2) relevant physical parameters of edges such as average luminance and contrast; (3) complex black-and-white images consisting of a multitude of edges with different orientations, average luminances, and contrasts; and (4) complex color images.

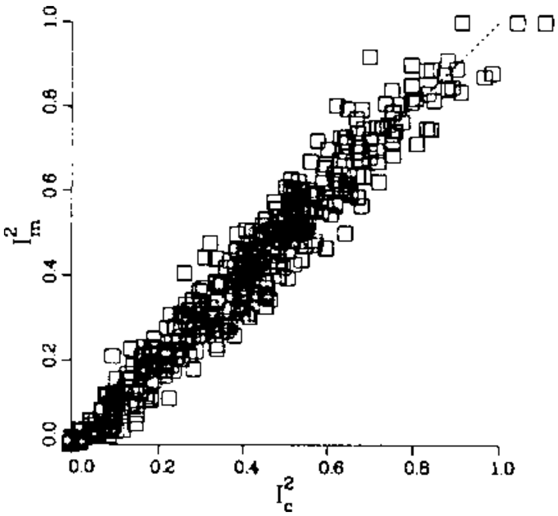

Figure 11. Squared experimentally measured category scaling data for perceptual impairments (I_m) versus the squared calculated perceptual impairments (I_c) of the complex color, the complex black-and-white, and the contrast and luminance experiment for the data averaged across subjects. For computational convenience, category scaling data were transformed linearly to fit the interval [0,1]. The dashed line indicates points wtih $I_c = I_m$.

Discussion

Because most of the current error measures are defined at the physical level, they are bound to neglect one or more of the perceptual aspects of the problem. Consequently, solutions are only optimal with respect to a perceptually suboptimal criterion and thus the solution will be suboptimal from a perceptual point of view. This leads to trial and error methods for defining useful physical error rneasures. Although this process may lead to a criterion that produces a perceptually satisfactory solution, it does not enable generalization.

Error measures defined at the physical level, such as the MSE, often neglect the fact that perceived image quality is a compromise between perceptual attributes instead of between physical parameters. In the specific case of sampling and interpolation, the physical error measures trade off distortion of edges and the distortion caused by periodic structure. Instead, the perceptual error measure compromises between the artifact's blur and periodic structure. If, for example, the number of edges in an image increases, the relative amount of the distortion at the edges increases likewise, whereas the overall impression of blur remains the same. Consequently, the perceptual error measure performs better in this case.

In Refs. 1, 5, and 6, we studied the performance of the perceptual error measure for sampled and interpolated images. A direct comparison of the experimentally measured impairments and the impairments predicted by the error measure can be found in Fig. 11. The figure summarizes the results for the data averaged over subjects for three experiments: the complex color (150 points), the complex black-and-white (108 points), and the contrast and luminance experiment (216 points). In these experiments, we used complex color images, complex black-and-white images, and the single-edged image, respectively (see Appendix). The experimental procedure was similar to the one described in the Appendix. All parameters of the model (σ_0, m_0, and λ_b) were unique for each subject and the parameter λ_b was also unique for each image (scene). The

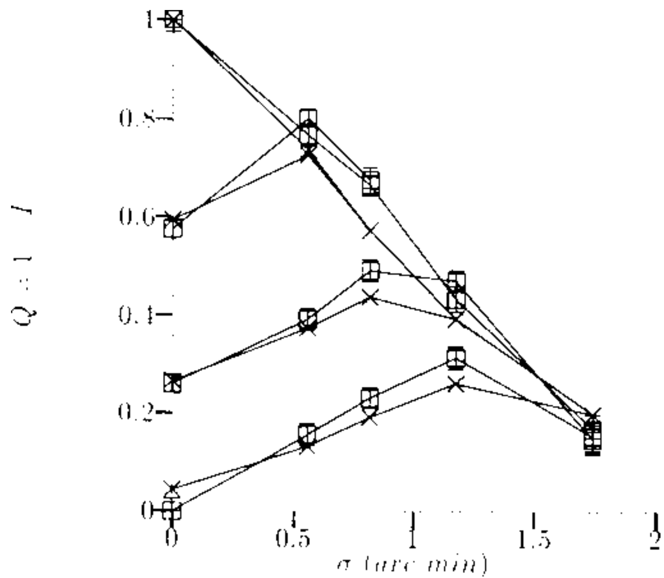

Figure 12. Comparison of experimentally measured category scaling data (□) and calculated data (×) for perceptual quality. The experimental perceptual quality data were averaged across subjects and plottted versus the spread parameter (σ) of the optical Gaussian interpolation filter for the complex "Terrasgeel" image of the complex black-and-white experiment. For clarity, category scaling data were transformed linearly to fit the interval [0,1]. The length of an error bar is twice the standard error of the mean. Often this length is smaller than the □ symbol. From top to bottom, the four pairs of curves are the nonsampled image and the sampled images with d = 1.18 arc min, d = 2.72 arc min, and d = 3.63 arc min.

Minkowski exponent was $\alpha = 2$. In Fig. 12, we present some of the experimental and predicted data of the complex black-and-white experiment, but now as a function of the spread parameter σ of the Gaussian interpolation filter.

From Figs. 11 and 12, we conclude that the perceptual error measure can adequately predict the perceptual impairment and thus image quality of sampled and interpolated images.

The Minkowski exponent $\alpha = 2$ adequately describes the data of individual subjects as well as the data averaged over subjects.[5] Other experiments indicate that the regularization parameter for blur λ_b is relatively independent of image content.[1,6] Since we included different types of images in these experiments, such as natural images, text, and abstract images in both black-and-white and color, we conclude that the model can also be applied successfully to other images. The λ_b parameter depends on the subject.[1,6] Typically, λ_b varies between 1 and 4 with an average value of 2 obtained for the data averaged across subjects. We stress the fact that such variations in the regularization parameter are not a shortcoming of the model but rather an explicit modeling of variations that exist among observers. Apart from variations in the low-level parameters σ_0 and m_0, the current model uses only one regularization parameter to account for the individual differences. The variations put a certain constraint on the optimization of image reproduction techniques, namely that optimal solutions should also possess a certain robustness against variations in λ_b. Graphs similar to the one in Fig. 12, but now for the data of individual subjects with extreme λ_b values, are shown in Ref. 1. These figures indicate that the model is also able to predict the data of individual subjects. Similar to Fig. 12, the variation in scaled perceptual quality across the data points is small. However, the absolute value is roughly enlarged by a factor that equals the square root of the number of subjects. Typical values for the intrinsic blur of the early visual pathway parameter and the threshold parameter for periodic structure are $\sigma_0 = 0.6$ arc min and $m_0 = 0.02$, respectively. While estimating these parameters from experimental data, we found a variation across subjects, which is not unusual for psychophysical experiments. Calculations[1] indicate that the influence of variations in the periodic structure threshold and intrinsic blur parameters on the regularization parameter for blur is only minor.

The perceptual error measure makes it possible to formulate a perceptually optimal combination of sampling and interpolation in perceptual terms. Since the relations between the sensorial strengths of the artifact blur and periodic structure and the physical parameters are known for a Gaussian interpolation function, the optimal physical parameters can be determined. In future work, similar relations for other interpolation functions may also be specified. Once these relations are deterrnined, we can immediately specify the optimal physical values. Note that the optimal solution of the sampling and interpolation problem in perceptual terms remains the same, irrespective of the physical implementation.

The same formalism can also be used to derive measures for more complex sampling and interpolation schemes as well as for other image processing techniques. In practical display systems, for instance, the electrical image signal can be processed before it is displayed. Although such a prefilter is unable to influence the perceptual strength of periodic structure, it can be used to reduce the strengths of the other artifacts. A prefilter can, for example, perform some deblurring. Unfortunately, such a filter itself introduces new artifacts, which must be included in the perceptual error measure.

Based on the results with sampling and interpolation, we think that the structure of the general image quality model is an adequate reflection of the perceptual process that leads to image quality. Hence, we expect it to work also for other image processing problems.

Acknowledgment. We are indebted to Dr. Huib de Ridder of the Institute for Perception Research for his help in setting up the experiments, analyzing, and interpreting the results and for sharing his knowledge of Minkowski metrics.

Appendix: Periodic Structure Experiment

Image. Only one type of image, namely a single-edged image, was used in this experiments This single-edged black-and-white image consists of two uniform regions with luminances L_1 and L_2 separated by a horizontal edge. The average luminance was $L = (L_1 + L_2)/2 = 9.6$ cd/m^2. The luminance contrast, defined as the Michelson contrast $C = (L_1 - L_2)/(L_1 + L_2)$, was $C = 0.21$.

Equipment. Single-edged black-and-white images were generated and displayed using a Gould DeAnza IP 8400 image processing system. An image consisted of 512 × 512 8-bit pixels. [The word *pixel* is restricted to indicate the elements of the combinations of the image processing system and the monitor used to display the stimuli. The area of a pixel is approximately equiluminant. Pixels lie next to each other and the distance between them (pixel pitch) is equal to the pixel size. Pixels are sufficiently small to ensure that artifacts such as periodic structure and "blockiness" are not visible. In addition, no moiré effects are introduced during simulation since sampling distances of the simulated sampling structures are multiples of the pixel pitch.] During the experiment, only the center 496 × 496 pixels (size 0.28 × 0.28 *m* or 4 × 4 deg) were shown on a Conrac model 2400 high-resolution 50 Hz interlace monochrome monitor. The viewing distance was 4 m and the pixel pitch was 0.49 arc min. The system was calibrated in such a way that there was a power-law relation $L \propto g^\gamma$, with an exponent $\gamma = 2.5$, between the 8-bit pixel value (g) and the luminance (L) of a pixel.

Stimuli. The stimulus set consists of different versions of the single-edged image, which was defined before. Besides this "original" image, the set contained single-edged images with a periodic structure varying in sampling distance and modulation depth of the structure. A vertical periodic structure was introduced by imposing a columnar sampling structure with a sampling distance of 3, 4, 6, and 8 pixel pitch units (1.47, 1.96, 2.94, and 3.92 arc min). The widths of the columns were 1, 2, 3, and 4 pixel pitch units (0.49, 0.98, 1.47, and 1.96 arc min), respectively. The modulation depth was varied by optical Gaussian filtering (see Eq. 6) in the horizontal direction. Gaussian filtering was simulated on the Gould DeAnza using a filter with a binomial impulse response $h(n)$ of length l[21]:

$$h\left(n-\left(\frac{l-1}{2}\right)\right)=\frac{1}{2^{l-1}}\frac{(l-1)!}{\left[(l-1)-n\right]!\,n!}. \tag{21}$$

The relation between the spread parameter (σ) of the Gaussian filter and the filter length (l) of the binomial filter was $\sigma = \frac{1}{2}\sqrt{l-1}$. For each sampling distance, modulation depth ranged front zero to maximal. Filter lengths were chosen such that the difference in perceived strength between successive filter lengths was approximately equal. The filter lengths are listed in Table I.

Thus far we have used the modulation depth of a periodic sine grating (Eq. 16). For the columnar periodic structure, we use the modulation depth of the first harmonic of the structure because the first harmonic is a good predictor for the visibility of any periodic structure of not too low a frequency[22]:

$$m = 2m_p \exp\left[-2\left(\frac{\pi}{d}\right)^2\left(\sigma^2+\sigma_0^2\right)\right]. \tag{22}$$

If this expression is used for the modulation depth, then the constant c in Eq. 15 must be changed in $c = (2/m_0)^{\beta 0}/\{[1 + (2/m_0)^3]^{1/3} - 1\}$. The factor m_p accounts for the additional interpolation of the columns of the structure. It can be shown that for columns with an ideal rectangular luminance profile, $m_p = \mathrm{sinc}(w/d) = \sin(\pi w/d)/(\pi w/d)$, where d is the sampling distance and w the width of the columns. Since in practice the profile will not be exactly rectangular, practical values will be lower than the theoretical values [sinc(1/3) ≈ 0.827 for the one-pixel-wide column and sinc(1/2) ≈ 0.637 for the other three widths]. Moreover, this additional attenuation may depend on the frequency and the luminance of the structure. We therefore measured the luminance profile of the columnar structure for different luminances and sampling distances with a Pritchard Photometer model 1380A combined with a spatial line scanner. The modulation depth of the first harmonic that fitted the profile best

TABLE I. Filter Lengths of the Binomial Impulse Response Filters of the Periodic Structure Experiment.*

Sampling distance (pixel pitch units)	Filter length (pixel pitch units)
3	1, 2, 3, 5, 6, 9
4	1, 3, 4, 6, 9, 12, 16
6	1, 2, 5, 8, 12, 17, 22, 29, 39
8	1, 6, 11, 17, 23, 31, 41, 53, 71

* If the filter. length is unity, then the filtered and unfiltered image are identical. Sampling distances are given for. a 4 m viewing distance, in which case the pixel pitch is 0.49 arc min.

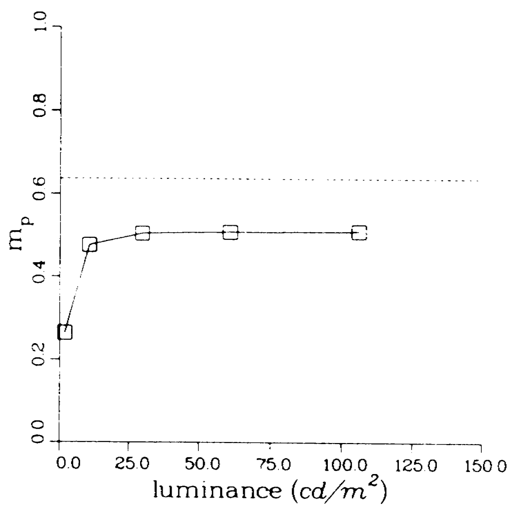

Figure 13. Additional attenuation of the modulation depth due to the width of the columns of the columnar structure wit hsampling distance d = 2.94 arc min versus the peak luminance of the luminance profile on the Conrac monitor. The column width was half the sampling distance. Hence the theoretical m_p value is $m_p \approx 0.637$. The theoretical value is indicated by the dashed line.

was used to calculate the value for m_p. Results are shown in Figs. 13 and 14. From the plots, we conclude that: (1) in practice, m_p values are lower than the theoretical values but the ratio of practical and theoretical values is to a good approximation independent of the width of the columns; (2) the value of m_p is to a good approximation independent of the sampling distance; and (3) the value of m_p is to a good approximation independent of the peak luminance of the profile for the luminances relevant to the experiment. For computational purposes, we use the m_p values of Fig. 14.

Procedure. Seven male subjects between 27 and 43 years of age participated in two sessions. Subjects had normal or corrected-to-normal vision and a visual acuity, measured on a Landolt chart between 1.25 and 2. Although two of the subjects had a slight red-green deficiency, their results did not differ significantly from those of the other subjects. Subjects rated both perceptual strength of the periodic structure and perceptual quality of the displayed images on a 10-point numerical category scale ranging from 1 to 10. Three subjects started with the quality session and four subjects rated perceptual strength of periodic structure first. Subjects received an instruction form in which the quality of the single-edged image was defined as depending only on the periodic structure in the uniform regions. Before the start of the actual experiment, subjects judged a test series of nine stimuli containing the

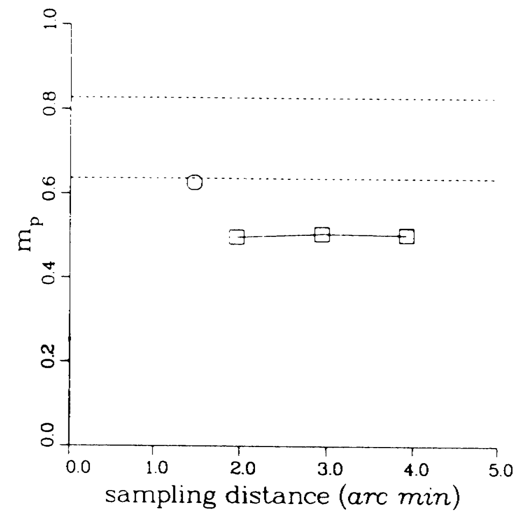

Figure 14. Additional attenuation of the modulation depth due to the width of the columns versus the sampling distance. (□) The peak luminance of the profile on the Conrac was 23 cd/m^2 and the column width (w) was half the sampling distance (d), hence the theoretical m_p value is $m_p \approx 0.637$. (○) The peak luminance was 35 cd/m^2 and $w = d/3$, so the theoretical value is $m_p \approx 0.827$. The theoretical values are indicated by the dashed lines.

extreme stimuli to adjust the sensitivity of their scale. All 32 stimuli were presented four times in each session except for the original image with zero modulation depth which was presented 20 times. The sequence of the images was random except that stimuli with the same sampling distance did not appear in consecutive trials. Images were presented for 5 s and followed by a homogeneous adaptation field with a luminance of 15 cd/m^2 that lasted until subjects pressed a key but had a minimum duration of 2 s. The viewing conditions satisfied CCIR Recommendation 500 (Ref. 25) except for the viewing distance, which was 4 m. All category data were transformed into an interval scale on the psychological continuum using Thurstone's law of categorical judgement. We applied a class I model involving replications over trials within a single sub jest with condition D constraints.[24] These constraints limit the number of model parameters by assuring that the correlation between the momentary position of stimuli and category boundaries as well as the dispersion of both category boundaries and stimuli are constant. Previous to the Thurstone correction, data were processed in accordance with Edwards' method[25] to correct scale values of the extreme categories. The Thurstone-corrected data were averaged across subjects. Averaging across subjects is allowed because, as shown in Fig. 15, the trends in the data are similar for all subjects. The group of subjects contained both subjects who were accustomed to doing scaling experiments as well as subjects who were not. In addition, two subjects knew the purpose of the experiment whereas the others did not.

References

1. M. R. M. Nijenhuis, Sampling and interpolation of static images: a perceptual view, PhD. Thesis Eindhoven Univ. of Technology, The Netherlands, ISBN 90-386-0412-2 (1993).
2. H. de Ridder, Subjective evaluation of scale-space image coding, *Human Vision, Visual Processing, and Digital Display II,* B. E. Rogowitz, M. H. Brill and J. P. Allebach, Eds., *SPIE Proc.* **1453,** 31–42 (1991).
3. J. Lubin, "A visual discrimination model for imaging system design and evaluation,"*Visual Models for Target Detection and Recognition*, Eli Peli, Ed., World Scientific Publishers, River Edge, NJ (1995).

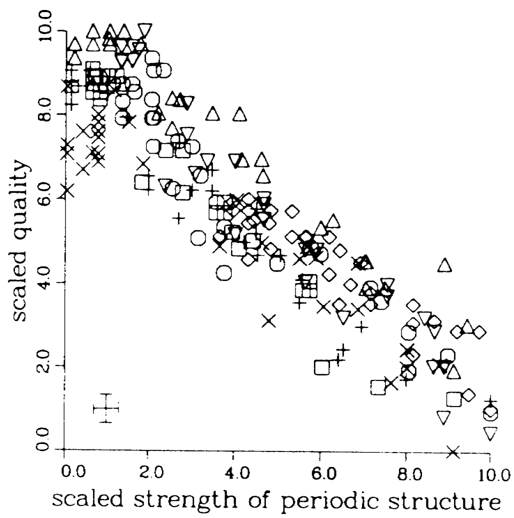

Figure 15. Experimentally determined category scaling data for perceptual image quality versus experimentally determined category scaling data for perceptual strength of periodic structure as measured in the periodic structure experiment. The symbols indicate data for different subjects. (□) FB, (○) GS, (Δ) HR, (+) MB, (∇) MN, (×) RS and (◊) SP. The error bar is twice the standard error of the mean averaged across the data points and subjects.

4. P. G. Engeldrum, "A framework for image quality models", *J. Imaging Sci. Tech* **39,** 312–318 (1995).
5. M. R. M. Nijenhuis and F. J. J. Blommaert, "A perceptual error measure and its application to sampled and interpolated single-edged images," *J. Opt. Soc. Amer. A* (in press).
6. M. R. M. Nijenhuis and F. J. J. Blommaert, "A perceptual error measure for sampled and interporated complex colour images," *Displays* **17,** 27–36 (1996).
7. H. de Ridder, "Minkowski-metrics as a combination rule for digital-image-coding impairments," *Human Vision, Visual Processing, and Digital Display III,* B. E. Rogowitz, Ed., *Proc. SPIE* **1616,** 16–26 (1992).
8. H. de Ridder, "Comparison of combination rules for digital-image-coding impairments," *IPO Annual Progress Report* **26,** 70–79 (1991).
9. J. H. D. M. Westerink, "Perceived sharpness in static and moving images", PhD Thesis, Eindhhoven Univ. of Technology, the Netherlands (1991).
10. W. S. Torgersoll, "Quantitative judgment scales," in *Psychological Scaling: Theory and Application,* H. Gulliksen and S. Messick, Eds., John Wiley & Sons, New York (1960), pp. 21–31.
11. D. W. Curtis and S. J. Rule, "Judgments of average lightness and darkness: a further consideration of inverse attributes," *Percept. Psychophy.* **24,** 343–348 (1978).
12. G. T. Fechner, *Elemente der Psychophysik I und II*, Breitkopf & Härtel, Leipzig (1860), p. 1907.
13. R. J. Watt, "Visual processing: computational, psychophysical and cognitive research," Erlbaum, Hove (1989).
14. H. R. Wilson, "A transducer function for threshold and suprathreshold human vision," *Biol. Cyber.* **38,** 171–178 (1980).
15. J. P. Oakley and M. J. Cunningham, "A function space model for digital image sampling and its application in image reconstruction", *Comput. Vision, Graphics, Image Process.* **49,** 171–197 (1990).
16. R. J. Watt and M. J. Morgan, "The recognition and representation of edge blur: evidence for spatial primitives in human vision," *Vision Research* **23,** 1465–1477 (1983).
17. G. E. Legge, "A power law for contrast discrimination," *Vision Research* **21,** 457–467 (1981).
18. C. R. Carlson and R. W. Cohen, "Visibility of displayed information: image descriptors for displays," Report ONR-CR213-120-4F, Office of Naval Research, Arlington, VA (1978).
19. A. Watanabe, T. Mori, S. Nagata, and K. Hiwatashi, "Spatial sine-wave responses of the human visual system," *Vision Research* **8,** 1245–1263 (1968).
20. T. Poggio, V. Torre, and C. Koch, "Computational vision and regularization theory," *Nature* **317,** 314–319 (1985).
21. J. B. Martens, "The Hermite Transform—Theory," *IEEE Trans. on Acoustics, Speech Sign. Process.* **38,** 1595–1606 (1990).
22. F W. Campbell and J.G. Robson, "Application of Fourier analysis to the visibility of gratings," *J. Physiol.* **197,** 551–566 (1968).
23. CCIR, "Method for the subjective assessment of the quality of television pictures," recommendations 500-3, in *Recommendations and reports of the CCIR*, International Telecommunication Union, Geneva (1986).
24. W. S. Torgerson, *Theory and methods of scaling*, John Wiley & Sons, New York (1958).
25. A. L. Edwards, *Techniques of attitude scale construction*, Appleton-Century-Crofts, New York (1957) .
26. H. de Ridder and G. M. M. Majoor, "Numerical category scaling: An efficient method for assessing digital image coding impairments," *Human Vision and Electronic Imaging: Models, Methods and Applications,* B. E. Rogowitz and J. P. Allebach, Eds., *Proc. SPIE* **1249,** 65–77 (1990).
27. J. H. D. M. Westerink and J. A. J. Roufs, "A local basis for perceptually relevant resolution measures," *SID 88 Digest of Technical Papers* **19,** 360–363 (1988).