Application of the general image-quality equation to aberrated imagery

Samuel T. Thurman[1,2,*] and James R. Fienup[1]

[1]The Institute of Optics, University of Rochester, Rochester, New York 14627, USA

[2]Currently with Lockheed Martin Coherent Technologies, 135 South Taylor Avenue,
Louisville, Colorado 80027, USA

*Corresponding author: sam.t.thurman@lmco.com

The regression analysis for the general image-quality equation (GIQE) [Appl. Opt. **36**, 8322–8328 (1997)] was performed using image data from well-corrected optical systems. We conducted human-subject experiments to examine the use of the GIQE with aberrated imagery. A modified image-quality equation for aberrated imagery is presented based on analysis of the experimental results. © 2010 Optical Society of America

OCIS codes: 110.3000, 110.3010, 110.4280, 280.4991.

1. Introduction

There are a number of different image-quality metrics [1]. Metrics that correlate well with the quality or utility perceived by a human observer are of particular interest here. For example, the Hotelling trace criterion, which measures the ability to perform a specific object-detection task, has been shown to exhibit a high correlation (correlation coefficient = 0.988) with the performance of human observers, who were asked to detect the presence of tumors in simulated medical images [2]. The square-root integral metric specifies subjective image quality in terms of *just noticeable differences* and also correlates well (correlation coefficient = 0.956 to 0.993) with data from human subjects [3]. The general image-quality equation (GIQE) [4], which specifies image quality in terms of the National Imagery Interpretability Rating Scale (NIIRS) [5], is central to this paper.

The NIIRS level of an image tracks the average ability of trained image analysts to perform a variety of object-detection, classification, and recognition tasks. Reference [5] lists example tasks for each NIIRS level. Larger NIIRS values indicate better image quality and the ability to perform more detail-oriented tasks. The scale is logarithmic by design, such that a change of ΔNIIRS $= 1$ is equivalent to a factor of 2 change in spatial resolution. A ΔNIIRS $= 0.1$ is considered to be barely noticeable by a human observer, while ΔNIIRS $= 0.2$ is easily perceived when comparing two images of similar quality [6,7]. The GIQE is used to predict the NIIRS level of imagery that can be expected from remote-sensing systems in various operational scenarios. Optical engineers can use the GIQE to design systems that meet desired NIIRS-level specifications. System operators can use the GIQE in tasking the appropriate system to fulfill customer data collection requests at specific NIIRS levels.

There are four versions of the GIQE. Versions 1.0 and 2.0 are not publicly available, but versions 3.0 and 4.0 are given in Refs. [4,5]. Both GIQE 3.0 and 4.0 have the form

$$\text{NIIRS} = c_0 + c_1 \log_{10}(\text{GSD}) + c_2 \log_{10}(\text{RER}) + c_3 \text{G}/\text{SNR} + c_4 \text{H}, \tag{1}$$

where GSD is the ground sample distance, RER is the relative edge response, G is the postprocessing

GIQE Version	c_0	c_1	c_2	c_3	c_4
3.0	11.81	−3.32	3.32	−1	−1.48
4.0 with RER ≥ 0.9	10.251	−3.32	1.559	−0.334	−0.656
4.0 with RER < 0.9	10.251	−3.16	2.817	−0.334	−0.656

[a]By convention, the GIQE is evaluated using GSD in units of inches.

noise gain, SNR is the signal-to-noise ratio, and H is the edge overshoot of an image. Table 1 lists the coefficient values for each version of the GIQE. Together, the GSD and RER terms represent a measure of spatial resolution; GSD is the detector-pixel resolution at the object, while RER tracks the impact of the system transfer function (including the effects of diffraction, aberrations, smear, and jitter) and any postprocessing (image sharpening) on resolution relative to the GSD. Based on the notion that ΔNIIRS = 1 is equivalent to a factor of 2 in resolution, one would expect the value of both c_1 and c_2 coefficients to be approximately equal to $\log_{10}(2) = 3.32$ [8]. The G/SNR and H terms account for the psychophysical impact of noise and edge-overshoot artifacts on the human visual system. Coefficients c_3 and c_4 are based on a regression analysis of data from images rated by trained image analysts. Table 2 contains some statistics of each of the GIQE terms for the imagery used in the development of GIQE 4.0 [4]. Additionally, this imagery was from well-corrected, conventional imaging systems. When compared with NIIRS ratings provided by trained image analysts, GIQE 4.0 exhibited a large coefficient of determination ($R^2 = 0.934$) for a set of test imagery.

Attempts to use the GIQE in scenarios not represented in the original regression analyses have failed. Reference [6] found GIQE 4.0 to be inaccurate in the low SNR regime (SNR ≤ 3). Additionally, GIQE 4.0 was found to be unsuccessful at predicting the NIIRS level of imagery from sparse-aperture systems [7]. Here, we considered the application of the GIQE to imagery from aberrated imaging systems in the case where the aberrations are known and compensated for with postprocessing. An experiment was conducted to determine ΔNIIRS image-quality variations in simulated imagery perceived by untrained human subjects. Results were compared with both GIQE 3.0 and 4.0. Further analysis led to a modified form of the GIQE that better fits the human-subject data, with a coefficient of determination $R^2 = 0.933$ to 0.945. It is worth noting that the results of this paper, as well as those of Refs. [6,7], were obtained using softcopy imagery, whereas GIQE 3.0 and 4.0 [4] were developed using hardcopy imagery.

Section 2 describes the design of the human-subject experiment. Section 3 contains experimental results for two sequences of imagery with different amounts of either (i) defocus aberration or (ii) mid-spatial-frequency wavefront errors. Section 4 contains an analysis of the results. Section 5 is a summary.

2. Experiment

This section describes the method of obtaining subjective ΔNIIRS values for aberrated imagery from human subjects.

A. Resolution Target

The experiment used simulated imagery of the tumbling E eye chart shown in Fig. 1. Strictly speaking, NIIRS measures the ability to perform specific object-recognition tasks listed in the NIIRS tables [4,5]. For the purpose of this study, however, image quality equates with the ability to recognize the orientation of the letter E at various scale sizes and contrast ratios. Thus, study participants did not need to be familiar with the tasks listed in the NIIRS tables. This resolution target has a number of other important features that enabled the use of untrained human subjects in the experiment, as opposed to using trained image analysts. The object-recognition task being

Table 2. Statistics of General Image-Quality Equation Terms for Imagery Used to Develop GIQE 4.0 [4]

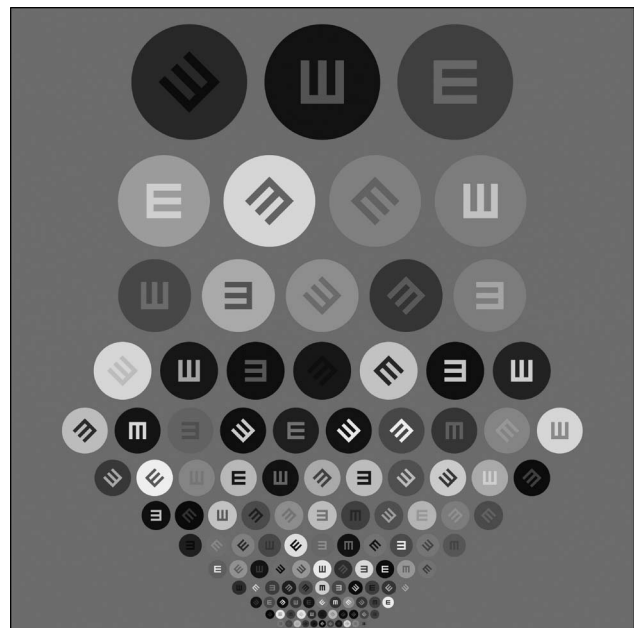| Term	Minimum	Mean	Maximum
GSD	0.076 m	0.52 m	2.03 m
RER	0.2	0.92	1.3
SNR	2	52.3	130
G	1	10.66	19
H	0.9	1.31	1.9

Fig. 1. Digital resolution target used for experiment.

performed does not vary with the scale of the image content. The organization of the eye chart allowed subjects to locate the minimum resolvable features in an image quickly, in comparison to scanning through a realistic scene for various-sized features. This allowed us to obtain more data from each subject in a given amount of time. The main drawback of using the eye chart is that it simply is not a natural scene. As such, there is no reason to believe that results with the eye chart will precisely match results that would have been obtained had we used actual aerial imagery. We had planned to use realistic synthetic imagery from the Digital Imaging and Remote-Sensing Image Generation (DIRSIG) model [9] in addition to using the eye chart, but did not have sufficient time to implement this.

The eye chart contains a wide range of feature sizes. The widths of the lines making up the letter Es in each row are equivalent to 2.00, 1.59, 1.26, 1.00, 0.794, 0.630, 0.500, 0.397, 0.315, 0.250, 0.198, 0.158, and 0.125 m on the ground (in order from the top to the bottom). This range allowed us to explore variations in image quality as large as ΔNIIRS $= \log_2(2/0.125) = 4.0$. Adjacent rows of the eye chart differ in scale by a factor of 1.26, which is equivalent to ΔNIIRS $= \log_2(1.26) = 0.33$. The eye chart was spatially oversampled to minimize artificial sampling artifacts when simulating images. The sampling of the eye chart was 0.042 m, which is a factor of 11 finer than the nominal GSD $= 0.46$ m of the simulated imagery.

Note that the eye chart exhibits a range of contrast ratios, instead of simply having black letters on a white background [8]. This was done to avoid skewing the results. For example, the impact of noise depends on the ratio of the noise amplitude to the contrast ratio of the feature of interest. Using only high-contrast-ratio Es would tend to minimize the impact of noise on the observed ΔNIIRS values, while using only low-contrast-ratio Es would make the observed ΔNIIRS values overly sensitive to noise. The background reflectance of the eye chart was set to 16%. The reflectances of the circular regions surrounding each E were randomly picked from a uniform distribution between 6% and 26%. The letter E reflectances were taken from the same random distribution, with the constraint that the absolute value of the reflectance difference between each E and the immediately surrounding circular region had to be greater than 1%. The linear gray scale for Fig. 1 (and all following simulated images) is such that black represents 5% reflectance and white represents 27% reflectance. While an antibunching method was used to balance the effect of orientation [10] in the resolution target, the effect of contrast was not ideally balanced. Specifically, note that the second and third rows of the eye chart do not contain very high contrast ratios. In retrospect, an antibunching method could have been used to yield a better balance of contrast in the target.

B. Baseline Imaging Parameters

Figure 2 shows simulated imagery of the resolution target for a system with the baseline imaging parameters listed in Table 3. These parameters are representative of a current high-resolution panchromatic remote-sensing system, such as WorldView-1 [11] or GeoEye-1 [12]. The image shown in Fig. 2(a) was obtained by (i) convolving the resolution target with a spectrally averaged optical point spread function (PSF) for the imaging system [13], (ii) convolving again with the pixel impulse response function for the focal plane array, (iii) downsampling this result based on the detector-pixel pitch p, (iv) scaling the image data to have an average number of detected photoelectrons per pixel N, and (v) adding shot and detector read noise. For (i), the spectrally averaged optical PSF was computed as a weighted sum of monochromatic optical PSFs computed at 81 discrete wavelengths using the gray-world (spatially spectrally separable) approximation for the scene. The
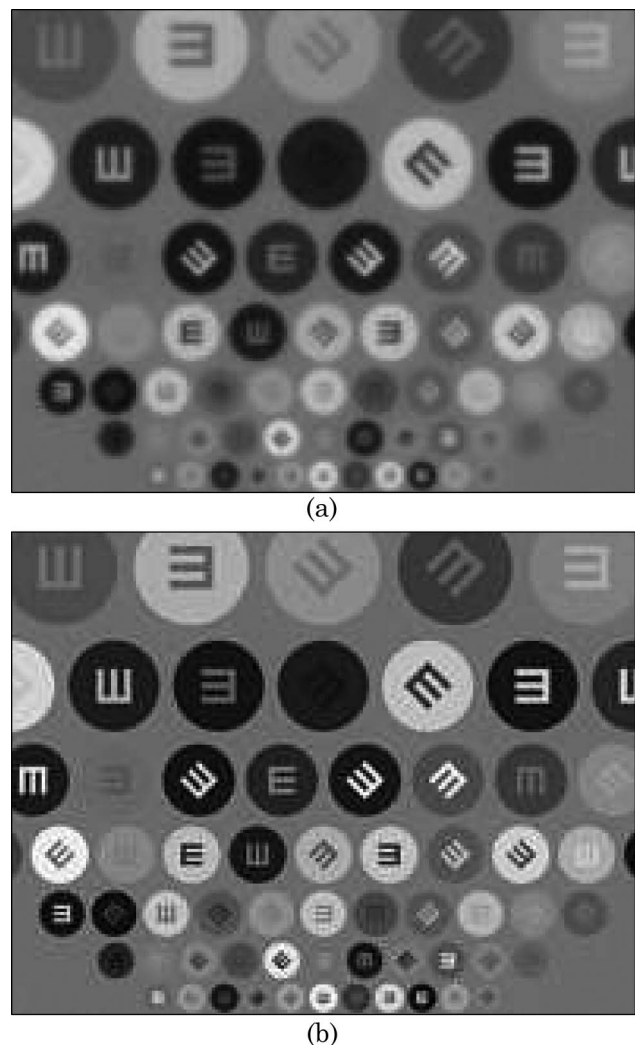

(a)

(b)

Fig. 2. Portions of the simulated baseline image (a) before and (b) after postprocessing. The portion shown here corresponds to rows 4–10 of the eye chart with corresponding line widths 1.00, 0.794, 0.630, 0.500, 0.397, 0.315, and 0.250 m (from top to bottom).

spectrum used for the scene was the product of a solar irradiance spectrum and a silicon-based detector responsivity curve. The GSD for this image is given by

$$\text{GSD} = pR/f, \tag{2}$$

where R is the altitude or target range of the system and f is the system focal length. The RER is calculated as [4,5]

$$\text{RER} = \text{ER}(0.5p) - \text{ER}(-0.5p), \tag{3}$$

where $\text{ER}(x)$ is the normalized edge response for the image, given by

$$\text{ER}(x) = \int_{0}^{\infty} dx' \int_{-\infty}^{\infty} dy \, s(x - x', y), \tag{4}$$

(x, y) are image plane coordinates, and $s(x, y)$ is the net PSF of an image, which includes effects from the optical PSF, the pixel impulse response, and an image-sharpening kernel (if the image was postprocessed). $\text{ER}(x)$ is also used to compute the edge-overshoot factor H as

as the signal amplitude for an 8% reflectance target divided by the average noise level (standard deviation) for an image before any postprocessing. Since the average reflectance of the resolution target is 16%, the signal amplitude for an 8% reflectance target is $N/2$, where N is the average number of detected photoelectrons per pixel. Thus, the SNR for an image is computed as

$$\text{SNR} = \frac{N}{2\sqrt{N + \sigma^2}}, \tag{6}$$

where σ is the standard deviation of the detector read noise. The postprocessed image shown in Fig. 2(b) was obtained by convolving the image of Fig. 2(a) with a symmetric 3 × 3 sharpening kernel of the form

$$w = \begin{bmatrix} -0.2 & -0.4 & -0.2 \\ -0.4 & 3.4 & -0.4 \\ -0.2 & -0.4 & -0.2 \end{bmatrix}. \tag{7}$$

The values of this kernel were chosen to yield an RER = 0.925 for the postprocessed image, which is about equal to the mean value of RER = 0.92 for the imagery used to develop GIQE 4.0 (see Table 2). The postprocessing noise gain G is given by

$$\text{H} = \begin{cases} \text{ER}(1.25p), & \text{when } \frac{d\text{ER}(x)}{dx} \geq 0 \quad \text{for all } x \in [1p, 3p] \\ \max\{\text{ER}(x)\} \text{ on interval } x \in [1p, 3p], & \text{otherwise} \end{cases}. \tag{5}$$

The H term acts as a penalty for edge-ringing artifacts. When using small (3 × 3 or 5 × 5), low-fidelity image-sharpening kernels, the net system transfer function is often boosted above unity at mid-spatial frequencies to achieve a desired net amount of image sharpening, resulting in significant edge-overshoot artifacts. The SNR term used in the GIQE is defined

$$\text{G} = \frac{\sqrt{\sum_{(m,n)} w_{m,n}^2}}{\sum_{(m,n)} w_{m,n}}, \tag{8}$$

where $w_{m,n}$ represents an individual element of w. Table 4 lists the individual image-quality terms for the baseline imagery, along with overall NIIRS-level predictions from GIQE 3.0 and 4.0.

C. Calibrated Reference Imagery

A set of reference images was simulated to have calibrated ΔNIIRS differences in image quality. The first approach we considered for doing this was to vary the SNR for the baseline imaging scenario. This method, however, was discarded for a number of reasons. The SNR coefficient in the GIQE is based largely on the psychophysical impact of noise on the human visual system. The general appearance of low-SNR images is considerably different than that of high-SNR images. The SNR coefficients for GIQE 3.0 and 4.0 do not agree. Finally, Ref. [6] found that

Table 3. Simulation Parameters for Baseline Image Shown in Fig. 2

Parameter	Value
Altitude	496 km
Primary mirror diameter, D	0.60 m
Obscuration diameter	0.23 m
Focal length, f	8.8 m
F-number, FN $= f/D$	14.7
Detector-pixel pitch, p	8.16 μm
Spectral bandwidth	0.4–1.1 μm
Mean wavelength, λ_0	0.644 μm
Detector sampling ratio, $Q = \lambda_0 \text{FN}/p$	1.19
Average number of photoelectrons per pixel, N	162,462
Detector read noise, σ	50 photoelectrons

Table 4. Image-Quality Parameters for Baseline Image

Parameter	Before Processing	After Processing
GSD	0.46 m	0.46 m
RER	0.459	0.925
SNR	200	200
G	1	3.42
H	0.876	1.057
GIQE 3.0 NIIRS	5.21	5.93
GIQE 4.0 NIIRS	4.75	5.32

GIQE 4.0 broke down for very low SNR values. The next approach for generating the reference imagery was to vary the primary mirror diameter D, which, in turn, affects the RER of the simulated imagery. Advantages of this approach are that it is deterministic and the resulting changes in image quality can be readily understood from the notion that each factor of 2 difference in resolution is equivalent to $\Delta \text{NIIRS} = 1$. The disadvantage of this approach is that GIQE 4.0 has a breakpoint at RER = 0.9. Thus, ΔNIIRS predictions from GIQE 3.0 and 4.0 for reference images generated by this technique would disagree. Ultimately, the method used to generate the reference imagery was to vary the GSD while maintaining a fixed RER = 0.925 for the imagery. In terms of fundamental system parameters, this could be accomplished either by changing the system altitude or simultaneously changing the detector-pixel pitch and primary mirror diameter. This approach is also deterministic and based on the notion that $\Delta \text{NIIRS} = 1$ is equivalent to a 2× change in resolution. Furthermore, both GIQE 3.0 and 4.0 (since RER > 0.9) predict the same variation of image quality for this scenario. This provides confidence in the quality of the reference images. In this scenario, the quality of the reference imagery varies as

$$\Delta \text{NIIRS} = \log_2(\text{GSD}_0/\text{GSD}), \qquad (9)$$

where $\text{GSD}_0 = 0.46\,\text{m}$ is the ground sample distance for the baseline image. A set of 38 reference images

Fig. 3. Portion of simulated reference image corresponding to $\Delta \text{NIIRS} = -0.98$ compared to the baseline image shown in Fig. 2(b).

were created, spanning a range of $\Delta \text{NIIRS} = +0.21$ to -2.38 in increments of $\Delta \text{NIIRS} = 0.07$. Figure 3 shows the $\Delta \text{NIIRS} = -0.98$ reference image for comparison with the baseline image shown in Fig. 2(b). The corresponding 2× difference in spatial resolution is apparent on comparison of the different rows of the eye chart in each image.

D. Subjective Image-Quality Evaluation

Observed ΔNIIRS values were obtained for simulated images from aberrated systems by having human subjects make visual comparisons with the calibrated reference image set using a MATLAB [14] graphical user interface (GUI). A number of precautions were taken to optimize this process. All images were displayed on a NEC MultiSync LCD2190UXi monitor calibrated to the IDEX-EPD profile using a GretagMacbeth Eye-One Display colorimeter and the NEC SpectraViewII calibration software. Subjects were permitted to adjust the display magnification to optimize the viewing eye scale for each image evaluation. 4 × 4 Lagrange interpolation [15] was used to resample each image for display at the monitor screen resolution. This was done to avoid image reinterpolation by either MATLAB or the display hardware. The GUI was designed to alternately display image pairs (one aberrated image and one reference image) in a softcopy flicker sequence. Image pairs were coregistered and displayed on the same absolute gray scale, and subjects could adjust the flicker rate to suit personal preference. The softcopy flicker sequence was deemed to be more comfortable for human subjects and likely to produce less eye strain than displaying images side by side. Finally, subjects could change the reference image being displayed in the flicker sequence.

For each aberrated image, subjects were instructed to choose the reference image that was a best match in terms of the ability to recognize the orientation of the Es in the eye chart. Because resolution, noise, and artifacts affect image quality differently, subjects were told to base their choice on the average ability to perform this task using the two smallest scale lines for which Es were recognizable in the aberrated imagery. The GUI presented the aberrated images in random order to each subject and recorded each subject's selections. The ΔNIIRS values of the selected reference images were then assigned to be the observed ΔNIIRS for the corresponding aberrated images. Six subjects (ages between 20 and 35) participated in the experiment, each of whom evaluated 56 images. While some of the subjects had prescription glasses, they were not required to wear them during the study.

3. Results

This section describes the imagery simulated with aberrations and the corresponding human-subject results.

A. Defocus Aberration

Twenty-eight test images were simulated with different defocus aberration amounts and SNR levels by modifying the parameters of the baseline imaging scenario. Defocus was included in the simulation as a rotationally symmetric, quadratic wavefront error in the pupil of the imaging system. The defocus amount was quantified by the peak-to-valley (P–V) amplitude (in units of the mean wavelength λ_0) of aberration across the area of the obscured pupil. Images were simulated with defocus amounts of $\{0, 0.125, 0.25, 0.375, 0.5, 0.627, 0.75\}$ waves P–V of defocus. The SNR of the imagery was modified by simply changing N, the average number of photons per pixel of the imagery. High, medium, and low 8% reflectivity SNR values of 200, 50, and 10 were obtained using $N = 162{,}462$, 12,071, and 1220 photoelectrons per pixel, respectively. All of the other imaging parameters were the same as those listed in Table 3. Figure 4(a) shows an example image with 0.5 wave P–V of defocus and SNR = 50 before postprocessing. Blurring due to defocus can be recognized by comparing this image with Fig. 2(a).

The postprocessing for the aberrated imagery did not use a simple 3×3 sharpening kernel, as was used for the reference imagery. Instead, a Fourier-domain Wiener filter [16–19] with more degrees of freedom was used to enable higher-fidelity aberration compensation. Also, the Wiener filter automatically yields postprocessed imagery that is optimum in terms of the expected mean-squared error without requiring a human observer "in the loop." Specifically, the Wiener filter used here was based on the formulation of Eq. (19) in [19], using filter parameters $c_n = 0.2$ and $c_a = 5$. The Wiener filter requires exact knowledge of the PSF (and thus, the aberrations) of the imaging system, as well as less-stringent statistical knowledge of both the object being imaged and the noise sources. Reference [19] gives a method of estimating from an image the power spectra for the object Φ_o and the noise Φ_n, where the noise is assumed to be white (Φ_n equals a constant) and Φ_o has the form

$$\Phi_o(\rho) = \begin{cases} A_0^2 & \text{for } \rho = 0 \\ A^2 \rho^{-2\alpha} & \text{for } \rho \neq 0 \end{cases}. \tag{10}$$

ρ is a Fourier-domain radial spatial-frequency coordinate and A_0, A, and α are power spectrum parameters. Here, however, the noise power spectrum is computed as

$$\Phi_n = N + \sigma^2, \tag{11}$$

where N is the average variance of the image shot noise and σ^2 is the variance of the detector read noise. For the object power spectrum, the value of A_0 was computed from knowledge of the average signal level as

$$A_0 = MN, \tag{12}$$

where M represents the number of pixels along each dimension of an image ($M = 313$ for each of the aberrated images in this study), and the values $A = 0.365 A_0$ and $\alpha = 1.57$ were predetermined by numerically fitting the power spectrum model of Eq. (10) to the fast Fourier transform (FFT) of the resolution target. Note that these expressions assume that unitary FFTs are used to compute Fourier-domain quantities and that ρ has units of Fourier-domain samples. Using these analytic expressions and predetermined values resulted in less postprocessing variability than if the power spectra parameters were estimated from the noisy imagery. Figure 4(b) shows the Wiener filter result for the aberrated image with 0.5 wave P–V of defocus and SNR = 50. The Wiener filter has improved the edge sharpness, but has also amplified the noise in the image.

A total of 28 images were simulated for this portion of the experiment: a set of seven images with SNR = 200 and each of the defocus values listed above, a set of seven with SNR = 50, and *two* sets of seven with SNR = 10. Figure 5 shows the variation of the image-
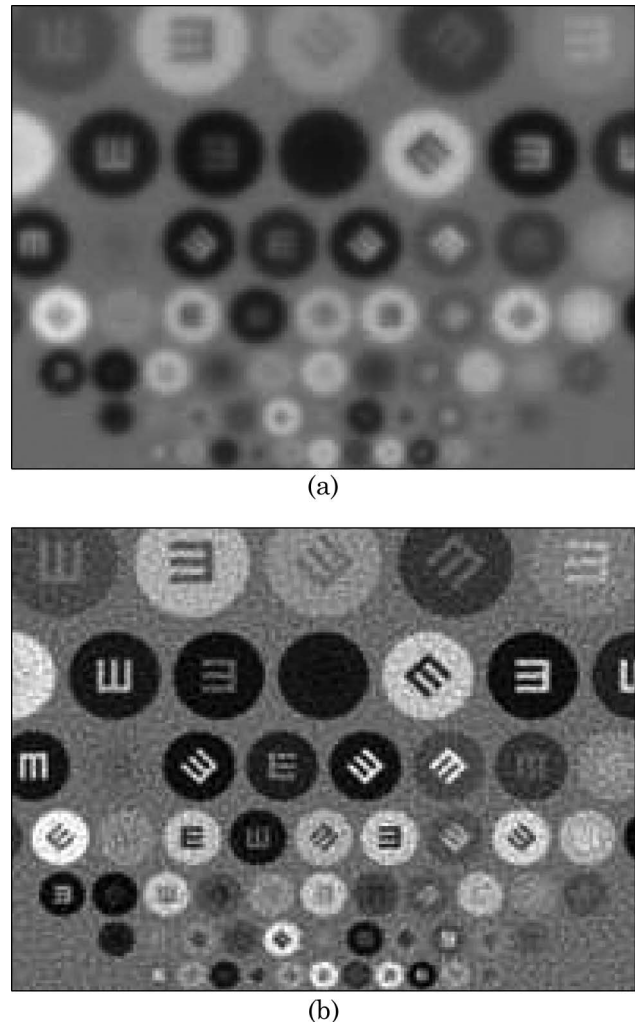

(a)

(b)

Fig. 4. Portion of simulated image with 0.5 waves P–V of defocus and SNR = 50 (a) before and (b) after postprocessing.
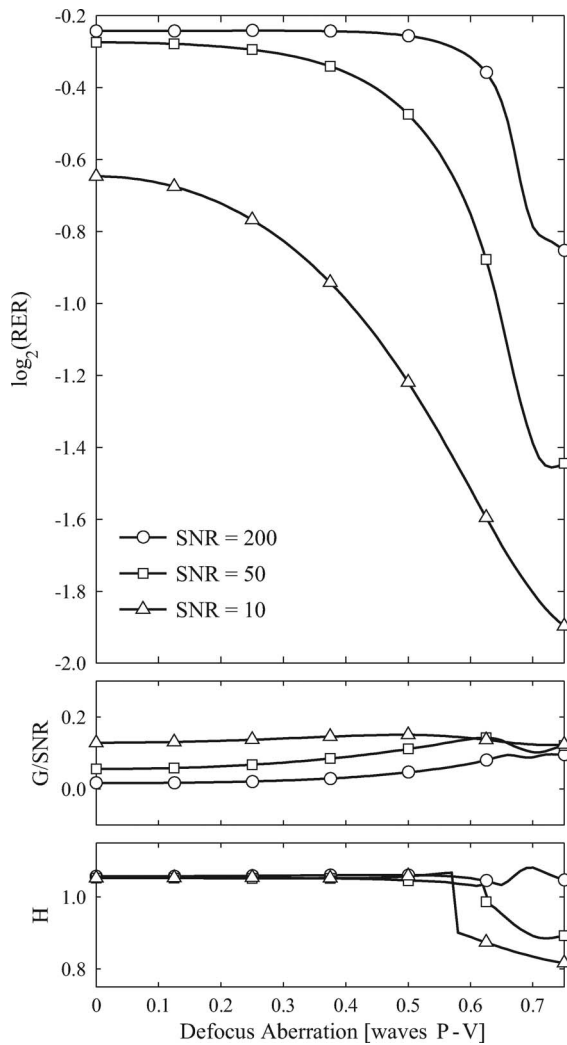
Fig. 5. Variation of $\log_2(RER)$, G/SNR, and H with the amount of defocus for SNR = 200, 50, and 10. The marked points represent the values for the images that were evaluated by human subjects.

quality terms $\log_2(RER)$, G/SNR, and H as a function of the amount of defocus for each SNR case. The marked points on each curve correspond to the values for the simulated imagery used in the experiment. Note that the vertical scale is equal for each plot. While the plots do not include the GIQE coefficients, it is obvious from the figure that RER is the dominant term affecting image quality. Figure 6 shows the observed ΔNIIRS values obtained from all six human subjects, along with ΔNIIRS values calculated from a modified GIQE (see Section 4). Note that many of the data markers in Fig. 6 represent multiple observations, because the ΔNIIRS values of the reference image set were quantized to increments of 0.07, and multiple observers often matched the same reference image to each aberrated image, especially for the SNR = 200 and 50 image sets. The spread in the observed ΔNIIRS values for these image sets are of the order of ±0.1. This seems reasonable, given that differences of ΔNIIRS = 0.1 are generally considered to be barely noticeable. One datum for the SNR = 50,

defocus = 0.75 wave P–V case was discarded as an outlier, because the observed ΔNIIRS value differed from the average by 0.7. Comparison of results from different subjects revealed that noise affected each subject differently, e.g., the observed ΔNIIRS values from some subjects where consistently either higher or lower than the mean for the low SNR = 10 imagery. This would explain the larger spread, of about ±0.2, in the ΔNIIRS values shown in Fig. 6 for the low SNR case. Section 4 contains a detailed analysis of the experimental results.

B. Mid-Spatial-Frequency Aberration

Several images were also simulated for a segmented-aperture system with mid-spatial-frequency figure errors on each segment. For this case, the system primary mirror was modeled as a ring of six hexagonal segments, each with a flat-to-flat diameter of 0.2 m, such that the overall diameter of the primary mirror was about equal to the value of $D = 0.6$ m used for the baseline imaging scenario (see Table 3). Figure 7 shows the spatial distribution of the mid-spatial-
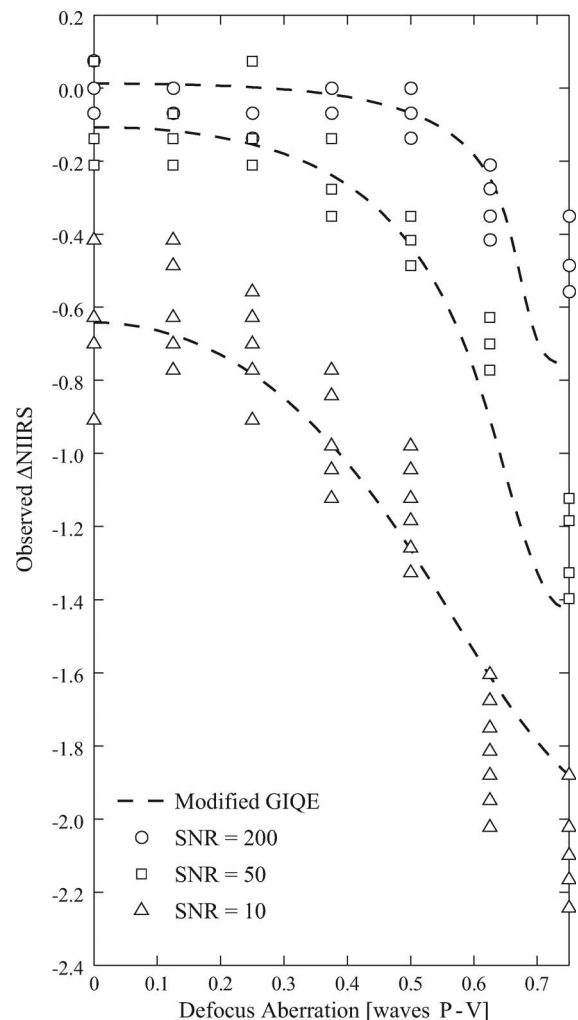

Fig. 6. Observed ΔNIIRS values obtained from human subjects for defocused imagery with SNR = 200, 50, and 10. The dashed curves represent the modified GIQE, ΔNIIRS = 0.291 + $\log_2(RER)$ − 2.229 G/SNR, obtained from the analysis in Section 4.
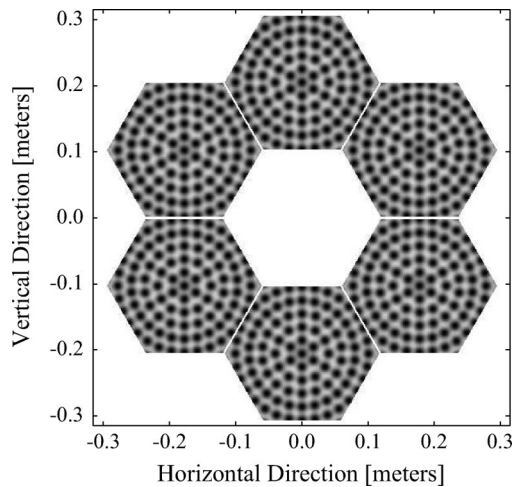
Fig. 7. Spatial distribution of mid-spatial-frequency aberrations for a segmented primary mirror.

frequency figure errors across the segmented primary. This particular aberration function is based on measurements (data supplied by NASA Marshall Space Flight Center) of the thermal deformation of the surface of a lightweight mirror [20]. The dominant spatial frequency of this aberration is 10 cycles per segment. The other fundamental system parameters (altitude, focal length, pixel pitch, spectral bandwidth, mean wavelength, and detector read noise) had the same values listed in Table 3 for the baseline imaging scenario. Figure 8(a) shows an example image simulated with 0.15 waves root-mean-squared (RMS) of mid-spatial-frequency aberration and SNR = 50 before postprocessing. While defocus aberration results in blurred imagery, the primary effect of a mid-spatial-frequency aberration is to reduce image contrast. This is a result of the optical PSF having a narrow central peak surrounded by a halo or intensity pedestal. As with the defocused imagery, a Wiener filter was used for postprocessing aberration compensation. Figure 8(b) shows the result of the Wiener filter for this example.

As for the defocus case, 28 images were simulated with different aberration amounts and SNRs: a set of seven images with mid-spatial-frequency aberration levels of $\{0, 0.125, 0.25, 0.375, 0.5, 0.627, 0.75\}$ waves RMS and SNR = 200, another set of seven with SNR = 50, and *two* sets of seven with SNR = 10. Figure 9 contains plots of $\log_2(\mathrm{RER})$, G/SNR, and H versus the aberration amount for each SNR case. Note that the RER for the segmented-aperture system is actually the geometric mean of the RER along the vertical and horizontal directions [4]. Likewise for the edge-overshoot term H. The observed ΔNIIRS values for these images along with ΔNIIRS values computed from a modified GIQE are shown in Fig. 10. Compared with the results for the defocused imagery shown in Fig. 6, there appears to be a bit more variability in the observed ΔNIIRS values for this case. All of the data for the SNR = 10, aberration = 0.3 waves RMS case were discarded, because this ima-

gery was matched to the worst reference image nine out of twelve times. Several subjects commented that they did this because the reference set did not include images with sufficiently poor quality for comparison with the imagery having SNR of 10 and 0.3 waves RMS aberration.

4. Analysis

The results of the human-subject experiment were compared to six different image-quality equations. Table 5 shows the analysis of the defocused-imagery results. The first column lists the image-quality equations that were considered. The first equation is essentially GIQE 3.0, while the second is GIQE 3.0 without the edge-overshoot term H. Similarly, the third and fourth equations represent GIQE 4.0 with and without H, respectively. The fifth equation represents the ΔNIIRS variation from $\log_2(\mathrm{RER})$ alone. The sixth equation contains a G/SNR term with a coefficient that provides the best agreement with the experimental data. The table lists
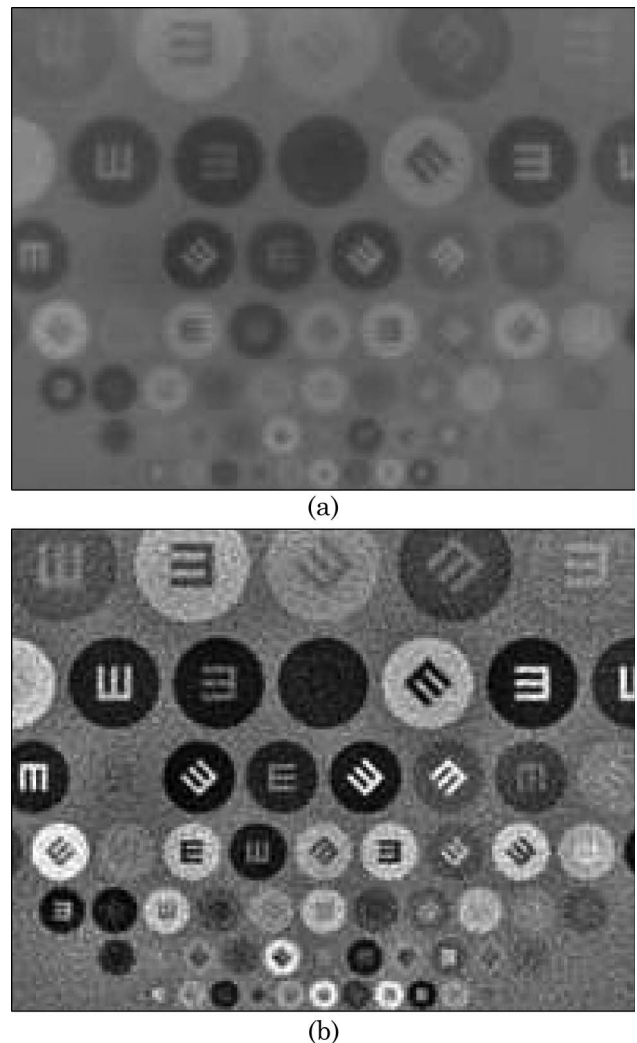

(a)

(b)

Fig. 8. Portion of simulated image with 0.15 waves RMS of mid-spatial-frequency aberration and SNR = 50 (a) before and (b) after postprocessing.
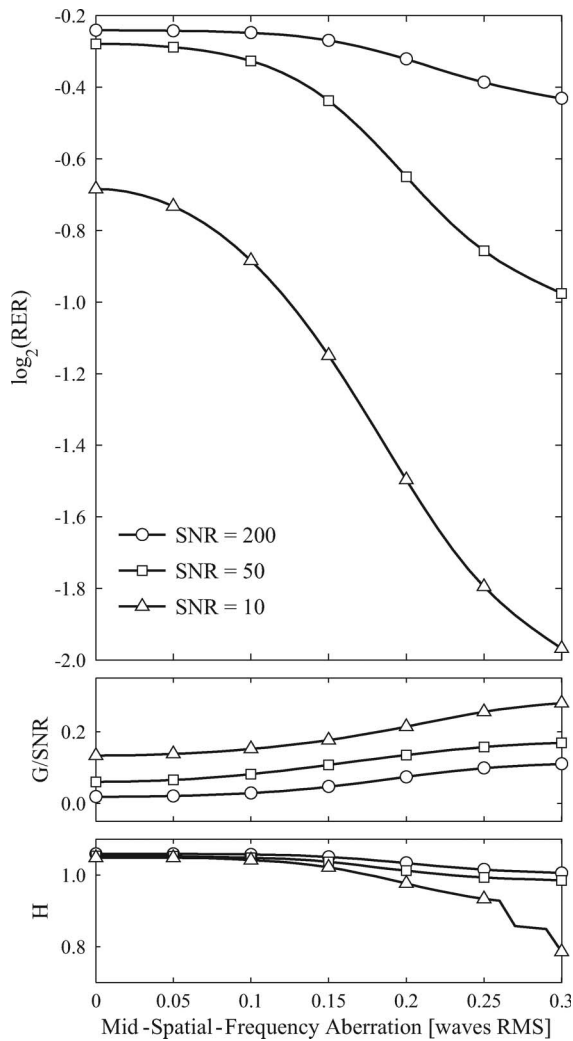
Fig. 9. Variation of $\log_2(\text{RER})$, G/SNR, and H with the amount of mid-spatial-frequency aberration for SNR = 200, 50, and 10. The marked points represent the values for the images that were evaluated by human subjects.

quantitative statistics that reflect how well each equation agrees with the experimental results. These statistics are the mean error μ_e, RMS error σ_e, and the coefficient of determination R^2, computed as

$$\mu_e = \frac{1}{K}\sum_k (\Delta\text{NIIRS}_{\text{obs},k} - \Delta\text{NIIRS}_{\text{iqe},k}), \qquad (13)$$

$$\sigma_e^2 = \frac{1}{K}\sum_k (\Delta\text{NIIRS}_{\text{obs},k} - \Delta\text{NIIRS}_{\text{iqe},k})^2, \qquad (14)$$

$$R^2 = 1 - \frac{\sum_k (\Delta\text{NIIRS}_{\text{obs},k} - \Delta\text{NIIRS}_{\text{iqe},k})^2}{\sum_k (\Delta\text{NIIRS}_{\text{obs},k} - \Delta\text{NIIRS}_{\text{obs,avg}})^2}, \qquad (15)$$

where $\Delta\text{NIIRS}_{\text{obs},k}$ represents an observed ΔNIIRS value from the experiment, $\Delta\text{NIIRS}_{\text{iqe},k}$ represents

a ΔNIIRS value predicted by an image-quality equation, $\Delta\text{NIIRS}_{\text{obs,avg}}$ is the average of the observed $\Delta\text{NIIRS}_{\text{obs},k}$ values, the subscript k is used to index various data points, and K is the number of data points used in computing each statistic. μ_e and σ_e were calculated individually for the SNR = 200, 50, and 10 datasets, as well as for the whole set of defocused imagery. The constant term for each image-quality equation was adjusted to yield zero mean error ($\mu_e = 0$) for the SNR = 200 imagery. This was done to highlight the fact that each equation fits the SNR = 200 data fairly well, with $\sigma_e \approx 0.1$, which is consistent with the ±0.1 variation in the observed ΔNIIRS values from different subjects for SNR = 200 (see Section 3.A). The equations differ, however, in how well they fit the lower SNR data. For example, the GIQE 4.0 equation had $\mu_e = 0.291$ for the SNR = 10 data, while the equation in the bottom row of Table 5 had $\mu_e = 0.040$. A large μ_e value (≥0.2) indicates a fitting error that is visually noticeable (given that image-quality differences of $\Delta\text{NIIRS} \geq 0.2$ are
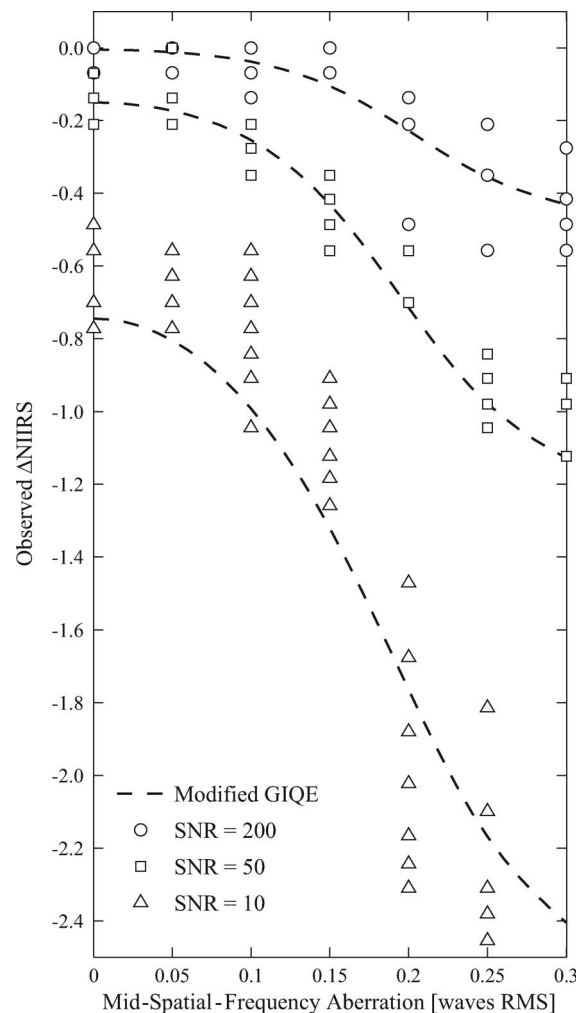

Fig. 10. Observed ΔNIIRS values obtained from human subjects for imagery with mid-spatial-frequency aberrations and SNR = 200, 50, and 10. The dashed curves represent the modified GIQE, $\Delta\text{NIIRS} = 0.283 + \log_2(\text{RER}) - 2.574\,\text{G/SNR}$, obtained from the analysis in Section 4.

Table 5. Analysis of Results for Imagery with Defocus Aberration

	SNR = 200		SNR = 50		SNR = 10		All SNRs			
ΔNIIRS Equation	μ_e	σ_e	μ_e	σ_e	μ_e	σ_e	μ_e	σ_e	R^2	Comment
$1.799 + \log_2(\text{RER}) - \text{G/SNR} - 1.48\,\text{H}$	0.000	0.108	0.072	0.119	0.244	0.369	0.140	0.274	0.812	GIQE 3.0
$0.238 + \log_2(\text{RER}) - \text{G/SNR}$	0.000	0.110	0.023	0.102	0.154	0.221	0.083	0.173	0.925	GIQE 3.0 w/o H term
$0.884 + 0.951\log_2(\text{RER})$ $- 0.344\,\text{G/SNR} - 0.656\,\text{H}$	0.000	0.096	0.085	0.127	0.291	0.366	0.168	0.271	0.815	GIQE 4.0
$0.192 + 0.951\log_2(\text{RER}) - 0.344\,\text{G/SNR}$	0.000	0.096	0.064	0.111	0.251	0.304	0.142	0.228	0.870	GIQE 4.0 w/o H term
$0.195 + \log_2(\text{RER})$	0.000	0.100	0.070	0.118	0.246	0.291	0.141	0.220	0.878	$\log_2(\text{RER})$ term only
$0.291 + \log_2(\text{RER}) - 2.229\,\text{G/SNR}$	0.000	0.130	−0.035	0.122	0.040	0.167	0.011	0.148	0.945	Fitted G/SNR term

considered to be readily perceived). The R^2 values in the second to last column indicate overall agreement with the experimental results.

The main points we wish to draw from the information in Table 5 are enumerated as follows.

1. Both GIQE 3.0 and 4.0 are inaccurate for the SNR = 10 imagery. The positive μ_e values indicate that the predicted ΔNIIRS values are overly optimistic, i.e., the GIQE predictions are consistently better than the observed image quality. The fact that $\mu_e > 0.2$ means that this discrepancy is at a visually noticeable level. This does not imply that either GIQE 3.0 or 4.0 is insufficient for predicting the quality of unaberrated imagery as they were originally intended to be used.

2. Dropping the H term from GIQE 3.0 and 4.0 yielded better agreement with the experimental results. An argument in support of this observation is that the Fourier-domain Wiener filter used here does not oversharpen the imagery, in the sense that the net modulation transfer function is boosted above unity. As such, any edge overshoot from the Wiener filter is associated with the Gibb's phenomena. Unlike the imagery used in the development of GIQE 4.0, which exhibited edge-overshoot values in the range of H = 0.9–1.9, the Wiener filter yields a maximum value of H = 1.03 [as explained in Ref. [8] and reflected by the values plotted in Fig. 5]. In this case, one might expect the H term to play a smaller role determining the image quality of Wiener-filtered imagery. For this reason, the H term was left out of the fitted image-quality equation on the bottom row of Table 5. Had a lower-fidelity Wiener filter been used for aberration compensation, due either to imperfect knowledge of the system transfer function or limited degrees of freedom, H might have played a more significant role.

3. The $\log_2(\text{RER})$ term alone agrees fairly well with the observed ΔNIIRS values. This is consistent with the notion that a ΔNIIRS = 1 is equivalent to a factor of 2 change in spatial resolution and the fact that RER is a dominant image-quality factor.

4. The equation on the bottom row of Table 5, with a $\log_2(\text{RER})$ term and a fitted G/SNR term, exhibited the best agreement with the experimental data with $R^2 = 0.945$. The G/SNR coefficient value of −2.229 was obtained by minimizing the RMS error σ_e for the entire set of defocused imagery. This value differs considerably from the G/SNR coefficients of GIQE 3.0 and 4.0. In general, aberrated imagery requires more aggressive postprocessing (more boosting of the mid and high spatial frequencies of an image) than unaberrated imagery. Also, a number of studies have shown that the impact of a fixed amount of noise on image interpretability depends on the spatial power spectrum of the noise [21–23]. From these two observations, it is not surprising that the G/SNR coefficient obtained here for aberrated imagery differs from the GIQE 3.0 and 4.0 coefficients obtained from a regression analysis of unaberrated imagery. Figure 6 shows ΔNIIRS values computed from this modified GIQE in comparison with the observed ΔNIIRS values. While the modified GIQE had the best overall fit, there are residual trends in the data that are not captured by this equation. For example, the ΔNIIRS predictions are a bit optimistic (by about 0.20 ΔNIIRS) for the SNR = 10 cases with more than 0.5 waves P–V of defocus aberration, as can be seen in Fig. 6. Perhaps the addition of a new GIQE term dependent on aberration strength would provide a better fit. The development

Table 6. Analysis of Results for Imagery with Mid-Spatial-Frequency Aberration

	SNR = 200		SNR = 50		SNR = 10		All SNRs			
ΔNIIRS Equation	μ_e	σ_e	μ_e	σ_e	μ_e	σ_e	μ_e	σ_e	R^2	Comment
$1.734 + \log_2(\text{RER}) - \text{G/SNR} - 1.48\,\text{H}$	0.000	0.115	0.078	0.130	0.198	0.394	0.112	0.283	0.835	GIQE 3.0
$0.195 + \log_2(\text{RER}) - \text{G/SNR}$	0.000	0.094	0.057	0.102	0.156	0.331	0.087	0.236	0.885	GIQE 3.0 w/o H term
$0.825 + 0.951\log_2(\text{RER})$ $- 0.344\,\text{G/SNR} - 0.656\,\text{H}$	0.000	0.122	0.113	0.163	0.295	0.458	0.167	0.328	0.777	GIQE 4.0
$0.143 + 0.951\log_2(\text{RER}) - 0.344\,\text{G/SNR}$	0.000	0.112	0.104	0.148	0.276	0.428	0.155	0.306	0.806	GIQE 4.0 w/o H term
$0.139 + \log_2(\text{RER})$	0.000	0.118	0.111	0.154	0.278	0.427	0.158	0.307	0.805	$\log_2(\text{RER})$ term only
$0.283 + \log_2(\text{RER}) - 2.574\,\text{G/SNR}$	0.000	0.075	−0.029	0.089	−0.036	0.250	−0.024	0.180	0.933	Fitted G/SNR term

of such a new term was beyond the scope of this paper. Nevertheless, it seems remarkable that the modified GIQE equation with a single fitted coefficient fits the observed ΔNIIRS values as well as it does, given the differences between this study and prior work on the GIQE noted earlier.

The corresponding analysis for the imagery with mid-spatial-frequency aberrations is summarized in Table 6. These statistics support the points drawn from the experiment with defocused imagery. Of particular interest is the similarity between the fitted G/SNR coefficient value of −2.574 for the mid-spatial-frequency aberration case to the value of −2.229 obtained for the defocus aberration case. This suggests that a G/SNR coefficient of approximately −2.3 might be more generally applicable to a wider range of aberration types. A larger study including diverse aberrations, however, is needed to substantiate this. Note that the geometric mean of the RER and H terms computed along the horizontal and vertical image dimensions should be used in the GIQE for asymmetric aberrations like coma and astigmatism (as was done for the imagery in Subsection 3.B).

5. Summary

We have described experimental methods for obtaining quantitative ΔNIIRS image-quality data from human subjects. Several details of doing this were addressed in order to avoid systematic errors and to facilitate the use of untrained subjects. This approach was then used to investigate the net loss in image quality due to aberrations at different SNR levels for the case in which the aberrations are known and compensated for with postprocessing. Both defocus and mid-spatial-frequency aberrations were considered. The results of this experiment were then compared with ΔNIIRS models based on the GIQE, which was originally developed for use with unaberrated imagery. Both GIQE version 3.0 and 4.0 were found to be inaccurate in this scenario for low SNRs. A modified image-quality equation was proposed for use with aberrated imagery as an alternative. This equation has a $\log_2(\text{RER})$ term, chosen based on the design of the NIIRS scale, and a G/SNR term with a coefficient fitted to the experimental results. The H term was omitted from the modified image-quality equation because the Wiener filter we used for aberration compensation does not result in large edge-overshoot artifacts. The G/SNR coefficients obtained for defocus and mid-spatial-frequency aberrations were −2.229 and −2.574, respectively. Overall, this modified image-quality equation agrees quite well with the experimental results, yielding coefficients of determination $R^2 = 0.945$ and 0.933 for the defocus and mid-spatial-frequency aberration cases, respectively.

References

1. For a broad, in-depth review of various image-quality metrics see H. H. Barrett and K. Myers, *Foundations of Image Science* (Wiley, 2004).
2. R. D. Fiete, H. H. Barret, W. E. Smith, and K. J. Myers, "Hotelling trace criterion and its correlation with human-observer performance," J. Opt. Soc. Am. A **4**, 945–953 (1987).
3. P. G. J. Barten, "Evaluation of subjective image quality with the square-root integral method," J. Opt. Soc. Am. A **7**, 2024–2031 (1990).
4. J. C. Leachtenauer, W. Malila, J. Irvine, L. Colburn, and N. Salvaggio, "General image-quality equation: GIQE," Appl. Opt. **36**, 8322–8328 (1997).
5. J. C. Leachtenauer and R. G. Driggers, *Surveillance and Reconnaissance Systems: Modeling and Performance Prediction* (Artech, 2001).
6. R. D. Fiete and T. Tantalo, "Comparison of SNR image quality metrics for remote sensing systems," Opt. Eng. **40**, 574–585 (2001).
7. R. D. Fiete, T. A. Tantalo, J. R. Calus, and J. A. Mooney, "Image quality of sparse-aperture designs for remote sensing," Opt. Eng. **41**, 1957–1969 (2002).
8. S. T. Thurman and J. R. Fienup, "Analysis of the general image quality equation," Proc. SPIE **6978**, 69780F (2008).
9. J. R. Schott, S. D. Brown, R. V. Raqueno, H. N. Gross, and G. Robinson, "An advanced synthetic image generation model and its application to multi/hyperspectral algorithm development," Can. J. Remote Sens. **25**, 99–111 (1999).
10. F. W. Campbell, J. J. Kulikowski, and J. Levinson, "The effect of orientation on the visual resolution of gratings," J. Physiol. **187**, 427–436 (1966).
11. http://www.digitalglobe.com/index.php/86/WorldView-1.
12. http://www.satimagingcorp.com/satellite-sensors/geoeye-1.html.
13. J. W. Goodman, *Introduction to Fourier Optics*, 3rd ed. (Roberts, 2005).
14. http://www.mathworks.com/products/matlab/.
15. A. K. Jain, *Fundamentals of Digital Image Processing* (Prentice Hall, 1989).
16. N. Wiener, *Extrapolation, Interpolation and Smoothing of Stationary Time Series* (Wiley, 1950).
17. H. W. Bode and C. E. Shannon, "A simplified derivation of linear least square smoothing and prediction theory," Proc. IRE **38**, 417–425 (1950).
18. C. W. Helstrom, "Image reconstruction by the method of least squares," J. Opt. Soc. Am. **57**, 297–303 (1967).
19. S. T. Thurman and J. R. Fienup, "Wiener reconstruction of undersampled imagery," J. Opt. Soc. Am. A **26**, 283–288 (2009).
20. J. B. Hadaway, P. Stahl, R. Eng, and B. Hogue, "Cryogenic test results of Hextek mirror," presented at Mirror Technology Days 2004, Huntsville, Alabama, USA, 17–19 August 2004.
21. P. F. Judy and R. G. Swennsson, "Lesion detection and signal-to-noise ratio in CT images," Med. Phys. **8**, 13–23 (1981).
22. P. A. Guignard, "A comparative method based on ROC analysis for the quantitation of observer performance in scintigraphy," Phys. Med. Biol. **27**, 1163–1176 (1982).
23. K. J. Myers, H. H. Barrett, M. C. Borgstrom, D. D. Patton, and G. W. Seeley, "Effect of noise correlation on detectability of disk signals in medical imaging," J. Opt. Soc. Am. A **2**, 1752–1759 (1985).