Up Periscope! Designing a new perceptual metric for imaging system performance

Andrew B. Watson; NASA Ames Research Center; Moffett Field, CA, USA

Abstract

Modern electronic imaging systems include optics, sensors, sampling, noise, processing, compression, transmission and display elements, and are viewed by the human eye. Many of these elements cannot be assessed by traditional imaging system metrics such as the MTF. More complex metrics such as NVTherm do address these elements, but do so largely through parametric adjustment of an MTF-like metric. The parameters are adjusted through subjective testing of human observers identifying specific targets in a set of standard images. We have designed a new metric that is based on a model of human visual pattern classification. In contrast to previous metrics, ours simulates the human observer identifying the standard targets. One application of this metric is to quantify performance of modern electronic periscope systems on submarines.

Background

The introduction of the electronic cameras and displays to the periscope viewing systems of the US Navy submarine fleet has brought with it an urgent need to characterize the image quality of the viewing system. Many other integrated imaging systems have a similar need for quality measurement. The quality must be characterized in performance terms, that is, in the degree to which it allows human crew members to perform specific relevant tasks. In the periscope case, an example task is the identification of watercraft.

Figure 1. Integrated imaging system.

Integrated imaging systems consist of assemblies of recording circumstances, optics, sensors, processing, transmission, display, and human viewing of image information (Figure 1). The last several decades have seen widespread introduction of integrated imaging systems into both public and private applications. Examples include commercial and government surveillance systems, air-, sea- and space-born military sensors, telepresence systems, drone cameras, as well as imaging systems in space and aviation. Introduction of these systems has brought with it a need to characterize the image quality of the integrated system. The quality must be characterized in performance terms, that is, in the degree to which it allows human users to perform specific relevant tasks. As an example, the introduction of electronic cameras and displays to the periscope viewing systems of the US Navy submarine fleet has highlighted the urgent need to characterize the image quality of the complete viewing system. In this case, an example task might be the identification of watercraft.

While it is possible to simulate many components of the imaging system to a desired degree of fidelity, to date it has not been possible to model the final component: the human observer. To complete the predictions of performance, it is necessary to model the visual performance of the human observer in a classification task. To accommodate all present and future imaging system effects, the inputs to the simulation of human performance must be images, rather than system parameters. In addition, the model must accurately simulate the performance of a human observer when the imagery is degraded by the imaging system.

Prior metrics

For the past 60 years, quantification of imaging system performance within the US military has been based on the Johnson metric [1-3]. In essence, this metric counts the number of visible sinusoidal cycles subtended by a target at the distance of interest, taking into account the contrast of the target and the human contrast sensitivity function. In a separate empirical procedure, human observers are tested to determine the number of cycles required for specific tasks. A serious shortcoming of these metrics is that human testing is very expensive, time-consuming, variable, and endless. A more fundamental problem with this approach is that it does not incorporate a model of human visual pattern classification. Instead it computes an ad hoc scalar measure of the "strength" of the average target image, and then empirically measures how much strength is required to perform a particular visual task. A characteristic feature of these metrics is that they are "parameter-based." While the human testing uses actual images, the calculation of metric values relies on a parametric description of the imaging system,

In recent years much work has been done to improve and extend the original Johnson method [4-6], and to apply these enhanced methods to a broad range of human target acquisition scenarios [7]. Among the enhancements are consideration of the
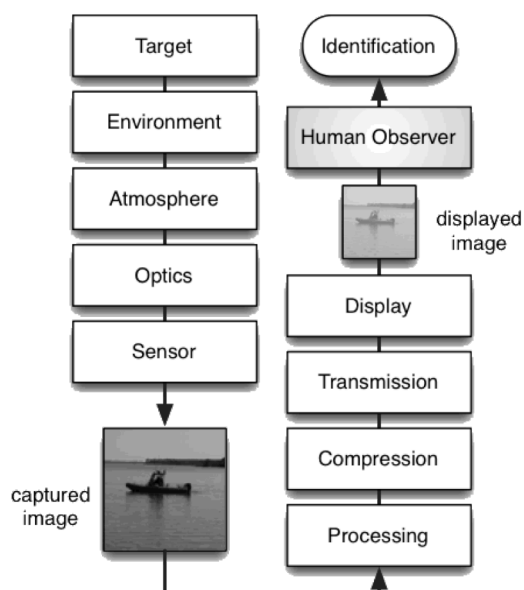
frequency spectrum of the target, the size of the target, and the spectrum and luminance dependence of noise in the imaging pipeline. These enhancements replace N with a more complex but predictive measure V. This enhanced approach and associated software packages (e.g., NVTherm, NVTherm2002, NVTherm-IP, NV-IPM, SSCAMIP, IICAM, IINVD, ACQUIRE0LC, and Detect05) are widely used in quantifying performance of military imaging systems [2, 7]. The US Army Night Vision and Electronic Signals Directorate (NVESD) has and continues to conduct extensive research on these metrics, and conducts extensive subjective testing to measure values of N_{50} or V_{50} for various classes of imagery [8-11].

This approach has been useful but still suffers from a number of practical and conceptual limitations.

1. It requires the use of human observers to establish the parameter N_{50} or V_{50} for a given task. Each new set of targets, and each new type of image artifact, requires new testing.
2. Because the approach is parametric and not image-based, it cannot accommodate new imaging system artifacts without modification and testing.
3. The approach does not incorporate a model of human target classification.
4. It quantifies identifiability of targets in terms of their average filtered energy, rather than the energy of the *differences* between targets in the set of candidates.
5. The approach does not predict the effect of magnification (displayed target size). Recently a modification to deal with magnification has been proposed [4], but it is not based on a human vision model.
6. It is unclear whether the parameter N_{50} or V_{50} as estimated empirically from targets degrade in one way (e.g., blur) generalizes to targets degraded by a completely different artifact (e.g. compression).
7. There is no single accepted standard for how to empirically measure N_{50} or V_{50}.

New approach

An alternative to the "parameter-based" approach is the "image-based" approach. In this approach, we physically simulate the system components that yield the displayed image. To complete the simulation, it is necessary to map the displayed image into human observer performance in target identification. That requires a realistic model of human visual pattern classification. The model must be "image-based" and must lead directly to identification performance when presented with samples from a small finite set of exemplars. It must also behave correctly in response to variations in contrast, size, noise, and other system artifacts. We have begun the development of an image-based human performance metric that is based on a previously developed model of human pattern classification [12].

Neural Image Classifier

The Neural Image Classifier (NIC) incorporates optical filtering, space-variant filtering and sampling by the midget retinal ganglion cells of the retina, neural noise, and ideal pattern classification. It is designed to predict the performance of a human observer attempting to classify samples from a finite set of images. Currently the model is achromatic, and incorporates time only by way of the duration of the target.

Target images

As noted, the model predicts classification of a finite set of images. This matches the scenario in which N_{50} is measured for Johnson-style metrics. Here we illustrate with an example of aircraft, as shown in Figure 2.

Optical Filtering

The target images are first blurred by a filter that simulates optical blurring by the human eye. This is accomplished using a formula recently proposed to describe the average human optical MTF for a given pupil diameter[13]. The pupil diameter is itself computed from a formula recently proposed to describe the average human pupil diameter under specified viewing conditions [14].

Neural Filtering

We then filter the images based on the action of the midget retinal ganglion cells (mRGC) of the human retina. These are the must numerous class of retinal ganglion cells, presumably responsible for spatial pattern vision. They represent a fundamental limit to transmission of visual information from eye to brain. The mRGC receptive field is modeled as a difference of Gaussians. The images are filtered by convolution with the mRGC receptive field, which varies in size with eccentricity. We determine the size at each eccentricity by means of a formula recently proposed for the density $d(\mathbf{x})$ of mRGC as function of position $\mathbf{x}$ in human retina[15]. To implement convolution by the mRGC kernel, we have made use of fast methods for space-variant filtering [16].

Noise, sampling, and eccentricity attenuation

We attribute the limiting noise to the output noise of the mRGC. In an image-based simulation, this requires that the noise variance be inversely proportional to the spatial density of the mRGC. As noted above, the mRGC density $d(\mathbf{x})$ decreases with eccentricity, and thus the noise increases. We model this by assuming a noise Power Spectral Density N at the foveal center, and instead of increasing noise with eccentricity, we attenuate image contrast by $1/d(\mathbf{x})$. We call this *eccentricity attenuation*.

Classification

Next we simulate an ideal classifier of signals known exactly. This consists of first computing the filtered templates corresponding to the set of target images. Then for each candidate image, we add noise and compute the match with each of the templates. The closest match is the selected as the classification result. Using fast methods [17] we can rapidly complete many such trials and estimate the confusion matrix, or the percent correct, to any desired degree of accuracy. We can also vary the contrast of the set of targets, and repeat the process, to generate a psychometric function, or to estimate a contrast that yields a particular percent correct.
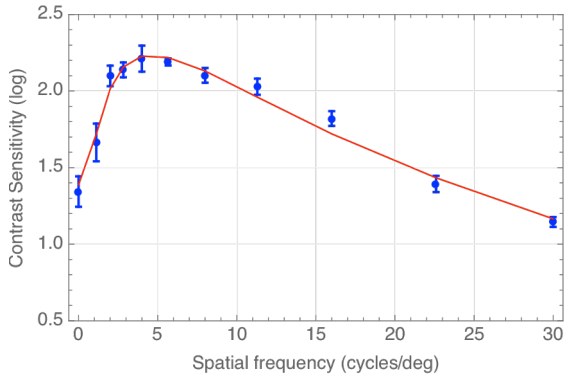

Figure 2. Aircraft images.

Figure 3. Contrast thresholds for Gabor functions and fit of NIC model.

Calibration

To generate predictions from the NIC it is necessary to estimate values for the key parameters of the model. These consist of the size of Gaussian center and surround of the foveal mRGC, the ratio of their weights, and the power spectral density of the noise (N). We have typically estimated these from contrast thresholds for a set of Gabor functions of fixed size (standard deviation = 0.5 deg) and frequencies of 0, 1.12, 2, 2.83, 4, 5.66, 8, 11.3, 16, 22.6, and 30 cycles/deg. An example set of calibration data is shown in Figure 3, along with the fit of the NIC model.

Application

Using the calibration predictions for an individual observer, we are now able to generate predictions for classification of images. As noted above, depending on the application, the images may be presented at various sizes, and one possible degradation is loss of contrast. As a preliminary test of the metric we have therefore measured contrast thresholds for classification of aircraft images that vary in size. Example results are shown in Figure 4, along with predictions of the NIC model. We express the thresholds in contrast difference energy [12]. This is the average contrast energy of the differences between each image and the mean image.
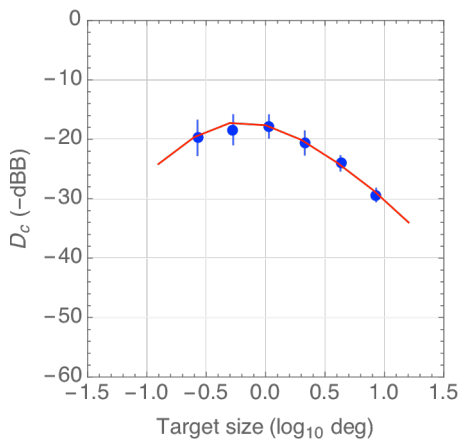

Figure 4. Contrast thresholds for identification of aircraft images and predictions of the NIC model. Contrast is expressed as contrast difference energy.

To register data and predictions, it was necessary to adjust the central efficiency of the classifier downward by a factor of 3 relative to the calibration results. This lower efficiency is consistent with previous findings that efficiency declines with pattern complexity [18]. Predictions that account for this effect await a satisfactory measure of complexity [19, 20]. However, we have observed that this efficiency is relatively constant over a range of target images of interest (aircraft, watercraft).

Building the metric

We have described the basic model of human image classification. To transform this into a practical metric of imaging system performance it is necessary to establish one or several sets of consensus standard images. Software simulations can then render the standard images at a desired size to simulate the action of an imaging system with a specified set of system parameters. The simulated images can then be classified by the NIC model, yielding a proportion correct (Figure 5). Alternatively, the NIC model may be used to determine the range at which the selected targets can be classified with a specified probability. We are currently engaged in these further steps in the development of an integrated imaging system performance metric.
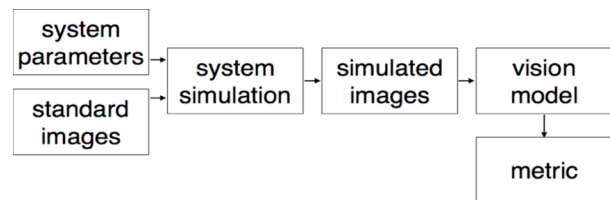

Figure 5. Elements of the integrated imaging system performance metric

References

[1]   J. Johnson, "Analysis of image forming systems," in *Image Intensifier Symposium, AD 220160*, ed: Warfare Electrical Engineering Department, U.S. Army Research and Development Laboratories, Ft. Belvoir, Va., 1958, pp. 244–273.

[2]   R. H. Vollmerhausen, E. Jacobs, and R. G. Driggers, "New metric for predicting target acquisition performance," *Optical Engineering,* vol. 43, pp. 2806-2818, 2004.

[3]   R. H. Vollmerhausen, D. Reago, and R. G. Driggers, *Analysis and evaluation of sampled imaging systems*. Bellingham, Wash.: SPIE Press, 2010.

[4]   B. P. Teaney and J. Fanning, "Effect of image magnification on target acquisition performance," *SPIE Proceedings,* vol. 6941, pp. 1-8, 2008.

[5]   J. D. Fanning, B. P. Teaney, J. P. Reynolds, and T. W. Du Bosq, "Range performance impact of noise for thermal system modeling," *SPIE Proceedings,* vol. 7300, pp. 730004-730004-12, 2009.

[6]   B. L. Preece, J. T. Olson, J. P. Reynolds, and J. D. Fanning, "Improved noise model for the US Army sensor performance metric," *SPIE Proceedings,* vol. 8014, pp. 801406-801406-11, 2011.

[7]   B. P. Teaney, T. W. Du Bosq, J. P. Reynolds, R. Thompson, S. Aghera, S. K. Moyer, *et al.*, "Human target acquisition

performance," *SPIE Proceedings,* vol. 8355, pp. 835510-835510-14, 2012.

[8] E. J. Boettcher, K. Krapels, R. Driggers, J. Garcia, C. Schuetz, J. Samluk*, et al.*, "Modeling passive millimeter wave imaging sensor performance for discriminating small watercraft," *Applied optics,* vol. 49, pp. E58-E66, 2010.

[9] R. G. Driggers, E. L. Jacobs, R. H. Vollmerhausen, B. O'Kane, M. Self, S. Moyer*, et al.*, "Current infrared target acquisition approach for military sensor design and wargaming," *SPIE Proceedings,* vol. 6207, pp. 620709-620709-17, 2006.

[10] K. Krapels, R. G. Driggers, D. Deaver, S. K. Moker, and J. Palmer, "Midwave infrared and visible sensor performance modeling: small craft identification discrimination criteria for maritime security," *Appl. Opt. ,* vol. 46, pp. 7345-7353, 2007.

[11] K. Krapels, R. G. Driggers, and J. F. Garcia III, "Performance of infrared systems in swimmer detection for maritime security," *Optics Express,* vol. 15, pp. 12296-12305, 2007.

[12] A. B. Watson and A. J. Ahumada, "Letter identification and the Neural Image Classifier," *Journal of Vision,* vol. 15, pp. 1-26, 2015.

[13] A. B. Watson, "A formula for the mean human optical modulation transfer function as a function of pupil size," *Journal of Vision,* vol. 13, pp. 1-11, May 31, 2013 2013.

[14] H. A. Peterson, A. J. Ahumada, Jr., and A. B. Watson, "Visibility of DCT quantization noise: spatial frequency summation," in *Society for Information Display International Symposium Digest of Technical Papers,*, Santa Ana, CA, 1994, pp. 704-707.

[15] A. B. Watson, "A formula for human retinal ganglion cell receptive field density as a function of visual field location," *Journal of Vision,* vol. 14, pp. 1-17, June 30, 2014 2014.

[16] J. S. Perry and W. S. Geisler, "Gaze-contingent real-time simulation of arbitrary visual fields," *Proceedings of the SPIE,* vol. 4662, pp. 57-69, 2002.

[17] A. B. Watson and A. J. Ahumada, Jr., "Predicting visual acuity from wavefront aberrations," *Journal of Vision,* vol. 8, pp. 1-19, April 22, 2008 2008.

[18] D. G. Pelli, C. W. Burns, B. Farell, and D. C. Moore-Page, "Feature detection and letter identification," *Vision Research,* vol. 46, pp. 4646-4674, 2006.

[19] A. B. Watson, "Perimetric complexity of binary digital images: Notes on calculation and relation to visual complexity," *Mathematica Journal,* vol. 14, 2012.

[20] A. B. Watson and A. J. Ahumada, "Modeling acuity for optotypes varying in complexity," *Journal of Vision,* vol. 12, pp. 1-19, 2012.

Author Biography

Andrew B. Watson received his PhD in Psychology from the University of Pennsylvania (1976), and did postdoctoral work at the University of Cambridge in England. He is the author of over 100 papers and seven patents on topics in vision science and imaging technology. He is the founder of the Journal of Vision and a Fellow of the Optical Society of America, the Association for Research in Vision and Ophthalmology, and of the Society for Information Display. In 2011, he received the Presidential Rank Award from the President of the United States.