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 users to perform specific relevant tasks. In the periscope case, an example task is the identification of watercraft.

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, INVD, 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 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 $d(x)$ of mRGC as function of position $x$ 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 $d(x)$ 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 $1/d(x)$. 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.](image-url)
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.

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.

References


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.