Modeling Visible Differences: The Computational Observer Model

Joyce E. Farrell[a], Haomiao Jiang[a], Jonathan Winawer[b], David H. Brainard[c] and Brian A. Wandell[a]
[a]Stanford University
[b]New York University
[c]University of Pennsylvania

Abstract

We describe a new approach to the design of visible difference metrics that is based on calculating the information available at different points in the visual pathway and determining whether an ideal linear classifier can use this information to reliably predict the difference between two stimuli. This approach, referred to as the computational observer model, establishes a near upper bound on performance by assessing whether the information necessary to make discrimination is present in the visual system. We describe the implementation of a software framework that embodies the computational observer and illustrate how this framework can be used to address specific questions about display design.

1. Introduction

One of the important goals of applied vision is to develop image quality metrics that engineers can use to optimize the design of information displays. The CIE color difference metric, CIELAB, is an excellent example of how applied vision research yielded a useful image quality metric that is widely used in industry today. Color matching functions and color difference metrics predict color discrimination performance for a standard observer. The standard observer model is based on visual psychophysical data collected from many individuals over many years. Watson and his colleagues extended the concept of the standard observer to develop metrics for spatial discriminations [1-3]. Zhang and Wandell [4] developed the SCIELAB metric to predict the detection and discrimination of visual stimuli that vary in both color and spatial dimensions. Visible difference metrics based on the standard observer concept have been extremely useful in both science and industry.

In this paper, we describe a new and complementary approach to designing visible difference metrics, which relies more fully on biology and computation. In this approach, we represent the information available at different points in the visual pathway, and we ask how well an ideal observer might perform with this information. The theory of an ideal observer has a rich history in both vision and engineering. Brindley wrote that the most secure method for relating perception and biology is to identify conditions in which observers cannot distinguish between stimuli, presumably because the stimuli produce the same biological response. He referred to psychological experiments that explore this principle as Class A experiments, and he included phenomena such as color-matching and threshold detection and discrimination [5, 6]. The concept of ideal observer theory also has roots in engineering [7] where a theory of ideal signal detection was developed. There have been many applications of the concept of the ideal observer in visual psychophysics, most extensively by W. S. Geisler's laboratory [8, 9].

To speed the development and experimentation with ideal observer metrics for industry, we built an open-source software system for estimating the biological representation of the visual stimulus at different stages of visual encoding. The system begins with a representation of the display image radiance; it then calculates the expected retinal irradiance after transformation by the cornea and lens. Next, it calculates the spatial array of cone absorptions after passing through the inert pigments and other properties of the cone photoreceptor mosaic. Finally, we have begun the process of implementing the expected pattern of retinal ganglion cell responses.

The basic premise of our approach is this: if two stimuli are statistically indiscriminable at one stage in this series of transformations, then the stimuli will be indistinguishable at all subsequent stages. For example, two stimuli with statistically indiscriminable radiance images when rendered on a display will be indistinguishable to a human observer. If two different radiance images pass through the cornea and lens and then produce indiscriminable retinal irradiance images, we expect the images will be indistinguishable to a human as well. Finally, two different irradiance images that produce statistically indiscriminable patterns of cone absorptions will be indistinguishable. To the extent that we can model subsequent stages in human image processing, such as retinal ganglion cell circuitry, we can continue to ask whether there exists a processing stage that renders the stimuli indistinguishable.

In many cases, however, it is not possible to specify an ideal observer analytically. In these cases, we can use the powerful tools of modern machine learning to approach ideal performance. We refer to this result as a computational observer model. We expect that performance of a well-designed computational observer will approach that of an ideal observer. Thus, when such a computational observer fails to detect the difference between two stimuli, we can reasonably predict that human subjects will also not be able to perform the discrimination. The converse, however, does not necessarily hold: if a computational observer successfully distinguishes a pair of stimuli, human observers may still fail to make the discrimination. In this way, computational observer metrics differ from standard observer metrics.

Computational and standard observer metrics are complementary. Standard observer metrics are based on phenomenological models whose aim is to predict performance under specific conditions. They are very useful because they summarize how well a typical subject will perform under the conditions covered by the model. The computational observer model approach we describe here asks a related but somewhat different question: Is there enough information in the visual system to detect the difference? For example, is there enough information in the spatial pattern of cone absorptions to detect the difference between two displays with different pixel structure and/or resolution? Is there enough information in the visual system so that a person might, under ideal conditions, discriminate the difference between two stimuli rendered on the same display using different methods? At what viewing distance will people fail to detect the difference between two rendered stimuli or two different displays? The computational observer model establishes an upper bound on performance by assessing whether the information necessary to make the discrimination is present in the visual system.

In this paper, we describe the implementation of a software framework that implements a computational observer. We illustrate how this framework can be used to address specific questions about display design.

 ISSN 0097-966X/14/4501-0352-$1.00

2. The Image Systems Engineering Toolbox for Biology (ISETBIO)

The Image Systems Engineering Toolbox for Biology (ISETBIO) is an open-source Matlab toolbox with data and functions that enable us to model the "front-end" of the human visual system [10]. ISETBIO includes code from several sources. It includes portions of the Image Systems Engineering Toolbox (ISET) developed by ImagEval Consulting, LLC to simulate the image processing pipeline for digital cameras [11, 12]. ISETBIO also includes the WavefrontOptics code developed by David Brainard, Heidi Hofer, and Brian Wandell to model human optics[13]. And it includes a set of tools implemented by Jon Winawer and Brian Wandell to transform cone absorptions into retinal ganglion cell spikes. ISETBIO and the Matlab script for the calculations we describe in this paper can be freely downloaded from the web [10, 13].

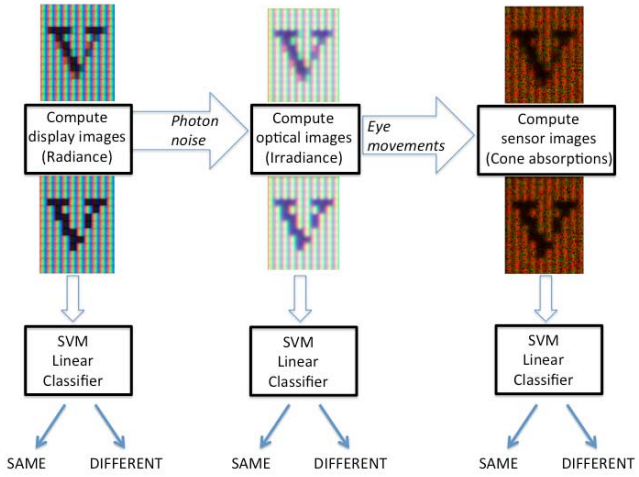

Figure 1. Visible difference metric based on ISETBIO and SVM classifiers. At each computational stage, we can determine the accuracy with which an ideal observer (modeled as a linear classifier) can reliably detect the difference between two stimuli.

Figure 1 shows how we use ISETBIO to build a visual difference metric. We model the early stages of human visual processing by characterizing the physical stimulus (display radiance image), the optical image impinging on the human retina (optical irradiance image), and the spectral, spatial and temporal sampling by the human cone photodetectors (sensor image). At each computational stage, we determine whether a computational observer (implemented as a set of linear classifiers) can reliably detect the difference between two stimuli. We use SVM to train a linear classifier and then calculate the performance accuracy of the linear classifier as a function of stimulus parameters.

2.1 Display radiance

We begin by characterizing the spatial-spectral radiance image that is the stimulus for the human eye. Assuming that the emissions from each pixel (and sub-pixel) are independent, the output radiance of a simulated display can be calculated as a linear combination of emitted spectra from each sub-pixel [14]. We simulate different rendering techniques (e.g. sub-pixel rendering methods such as ClearType) and calculate the spatial-spectral radiance of the displayed image. The radiance map for stimuli are represented as ISET scene structures that are the input for the next stage of image processing – the human optics.

2.2 Retinal irradiance

At the level of human optics, we take the radiance map as input and compute the corresponding irradiance optical image. Computation in this level is based on the WaveFrontOptics code [13] and data provided by Thibos [15]. We can use the data to build models of different types of eyes and vary parameters such as pupil size, eye length, focal length and the effects of aging. We compute the optical image by convolution with a point-spread function (PSF) that is calculated for different individuals. Point-spread functions with different degrees of defocus and astigmatism are shown in Figure 2 below. The output of this layer is an irradiance map stored in an OI structure, which can be used as input to a cone absorption layer.

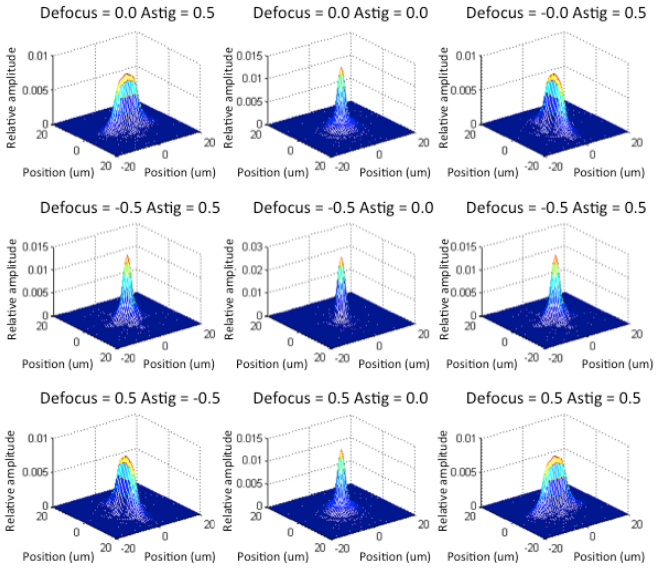

Figure 2. Point spread functions varying in defocus and astigmatism.

2.3 Cone absorptions

We convert the retinal irradiance into photoreceptor absorptions using the Stockman-Sharp functions, which are maintained on the very thorough and useful CVRL site[16]. These functions include estimates of the ocular transmittance and cone photopigment absorption functions that are needed to calculate how to convert retinal irradiance into the expected rate of photon absorptions in each of the three different cone mosaics. The implementations in ISETBIO and the Psychtoolbox [17] are in good agreement with one another.

The standard human cone fundamentals include assumptions about the ocular transmittance (macular pigment and lens), as well as the photopigment absorption functions. To enable us to perform analyses of the effects of these separate components, the software represents each of these components, as well as density and related parameters that can be set to characterize different populations. For example, smokers have lower density macular pigment than non-smokers [18].

The effective cone absorptance is the product of the ocular transmittance and cone photopigment absorptance. A number of additional parameters can also be specified, including the cone spatial density, field of view, and number of cones from each

class. We simulate dichromacy by setting the percentage of receptors in one of the cone classes to zero,

Finally, we typically estimate the number of cone absorptions using a ten-millisecond temporal sampling rate. Visual perception typically integrates absorptions for tens of milliseconds, but we use the shorter sampling rate so that we can account for the effects of small eye movements.

2.4 Eye movements

Small saccadic eye movements are modeled as a linear drift in position over time plus a two-dimensional Gaussian noise term.

$$(xy) = (\beta x \beta y)t + \beta 0 + \varepsilon,\ \varepsilon \sim N(0,\Sigma)$$

In the equation above, the first term corresponds to the eye drifting while the noise term corresponds to microsaccades. By fitting this linear equation to eye movements measured during short trials (50 ms), we find the covariance matrix of microsaccades and use this to randomize the position of the eye at each sample point for the computational observer model.

2.5 Statistical discriminability of the signal

The ideal observer is a statistical decision maker that performs as well as possible at classifying between different stimulus types. Ideal observer formulae can be derived for certain classes of stimuli and stages of visual processing, but the ideal observer is not known for every condition of interest.

As an approximation to the ideal observer, we use support vector machines (SVM). Specifically, we train a support vector machine (SVM) to assess how accurately the biological response patterns of two types of stimuli can be discriminated. For example, we can train an SVM to classify based on the retinal irradiance of two stimuli, which is subject only to the uncertainty of the photon noise. We can then train an SVM to classify between these same stimuli after they have transformed into cone photoreceptor absorptions, assuming no eye movements. We can then ask how well the stimuli can be discriminated in the presence of eye movements, and so forth.

We use the regularized support vector classification (linear kernel) algorithm implemented in the widely distributed libsvm package [19]. Other methods may be appropriate and might be usefully compared, such as Accuracy Maximization Analysis [20-21].

3.0 Predictions

We simulated an observer with a 3mm pupil diameter and an optical point spread function for an average human eye with no defocus or astigmatism (see Figure 2). The optical aperture of each cone photoreceptor was set to 1.5 microns (diameter), and we experimented with different ratios of L,M and S cones. The modeled stimuli occupy 1 deg of the visual field, but we only used the cone absorptions in the central 0.2 x 0.2 deg to make the predictions.

3.1 Contrast sensitivity functions

Using ISETBIO, we calculated the ability of a computational observer to detect luminance contrast gratings for a range of spatial frequencies. The stimuli were vertical gratings multiplied by a Gaussian envelope (Gabor patch).

To analyze a single stimulus condition, say one spatial frequency at one contrast level, we simulated 6000 trials, half with gratings and half with uniform fields; each trial represents the cone absorptions over a 50 ms period. These absorptions were calculated by summing the cone absorptions from five different eye positions separated by 10 ms intervals. Hence, the cone absorptions for each simulated 50 ms trial was based on 5 different eye position samples.

We trained the SVM classifier to discriminate gratings from uniform stimuli using 10-fold cross-validation. To accomplish this, we grouped simulated cone absorptions into 10 sets of 600. Each set contains 300 grating and 300 uniform simulated cone absorptions. We held out one set of stimuli, and trained the classifier using the other nine. We measured classification performance on the held-out stimuli, repeating this process 10 times to estimate classification accuracy.

The cone absorptions were calculated in the presence of photon noise and eye-movements and we calculated the classification accuracy as a function of grating contrast. Figure 2 shows the classification accuracy for a computational observer with half as many L-cones as M-cones. We also calculated classification accuracy functions for a computational observer with twice as many L-cones as M-cones. For both model observers, we estimated the contrast level required for 80% accuracy. We plot the contrast sensitivity (the inverse of the 80% correct contrast level) for a range of grating frequencies in Figure 3.

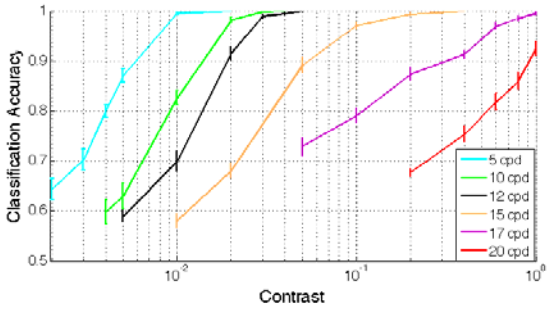

Figure 2. Classification accuracy plotted as a function of grating contrast with spatial frequency as the parameter.

The contrast sensitivity function for these conditions falls off with spatial frequency. The information in the cone mosaic is quite good even at the lowest spatial frequencies. Hence, as Schade pointed out in his seminal work in 1956 [23, 24], the measured loss of sensitivity in the low spatial frequency regime must be of neural origin.

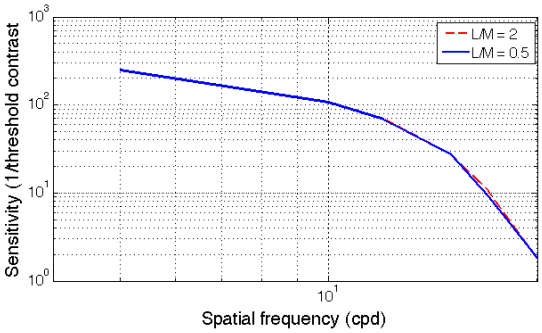

Figure 3. Contrast sensitivity functions for computational observers with two different L/M cone ratios.

We also note that the computational observer's contrast sensitivity functions do not change as we altered the L/M cone ratio (Figure 3). Gunther and Dobkins [25] showed experimentally that human contrast sensitivity is invariant to L/M ratio. The computations and empirical measurements suggest that the visual system retains its spatial acuity despite individual differences in L/M cone ratios.

3.2 Letter discrimination

In a second analysis, we calculated the ability of a computational observer to discriminate between binary and ClearType versions of the same letter (see Figure 1). The analysis was carried out to understand the impact of ClearType as display resolution increases. In one experiment, the letters were rendered on a simulated low-resolution display (72 pixels/inch), modeled after a Dell LCD (Model Number 1905FP) display [ref]. In the second condition the letters were rendered on a simulated high-resolution (300 ppi) theoretical prototype of this display. The ClearType letters were created using a three-tap filter defined by a single parameter (a = 0.2) [ref].

The two letters ('g" and "s") were taken from the Microsoft Georgia 10 point font. The letters were simulated to be the same size on both the 72 and 300 ppi displays. The size of the retinal irradiance image was a function of the letter size and viewing distance. When the viewing distance was 0.6 meters, the letters subtended a visual angle of 0.3371 degrees.

For each letter and display resolution, we simulated 1500 trials, half with binary versions and half with ClearType versions of the same letter. As in the previous analysis, we trained the SVM classifier to discriminate between the binary and ClearType version of the letters using 10-fold validation.

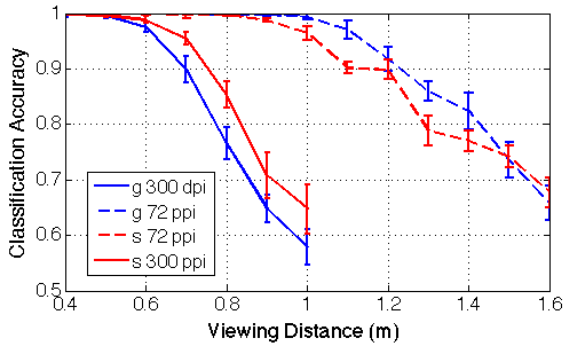

Figure 4. Classification accuracy for binary versus ClearType renderings plotted as a function of viewing distance.

The computational observer is always able to detect the difference between binary and ClearType versions of the same letter at 0.4 m (Figure 4). On a 300 dpi monitor, classification accuracy decreases to less than 70% at a viewing distance of 0.9m. On the 72 ppi display the ClearType rendering reaching 70% accuracy at a much greater viewing distance (1.6m). These calculations show that at today's resolution and viewing distances the average cone mosaic contains enough information to enable people to see the difference between binary and ClearType font renderings.

4.0 Conclusions

Standard observer models have provided a useful approach to visible difference metrics in both science and industry for nearly a century. The standard observer model is useful as a phenomenological description of typical observer performance for a specific task. Typically, standard observers are defined for a class of experiments, such as color or pattern. The observer is defined by a small set of data and key formulae.

The computational observer model described here uses the power of modern computing and new advances in our knowledge of optics and retinal encoding to ask a broader question: Is the information necessary to detect a visible difference - any difference - between two stimuli present in the visual response at well-specified stages of visual processing? We have found the computational observer model to be helpful in several ways. For scientific investigation, it is useful to understand what information is present at different points in the visual pathway and further to understand how information is lost or preserved. Further, the goal of building a computational observer model is helpful in guiding the selection of biological and human experiments. This line of work will be described in future publications.

For engineering applications, the computational observer model is helpful in at least two ways. First, it represents a reasonable upper bound on human performance. When the computational observer cannot make a reliable discrimination, we expect that the discrimination will be beyond the capability of the human observer as well. This provides a useful guide to prevent unnecessary costs that would be required to build devices to a specification that exceeds the limits of human perception. Second, as the ClearType example illustrates, the computational observer can be tested using arbitrary visual stimuli seen on theoretical prototypes. The computational observer is a unified model that integrates metrics for pattern, color and time; it is structured to permit us to incorporate individual differences that span the range of the typical population (L/M ratios, wavefront aberrations, etc.). This makes the computational observer a useful tool for display design.

5. Acknowledgements

The authors wish to thank Michael Bennett and Shalomi Eldar for their work on the Matlab software for rendering ClearType fonts on simulated displays. We also thank Kevin Larson, Tanya Matskewich, and Greg Hitchcock for their advice and encouragement. And we thank the Microsoft Corporation for supporting this research.

6. References

[1] A. B. Watson, "The Spatial Standard Observer," *Journal of Vision,* vol. 4, 2004.

[2] A. B. Watson and A. J. Ahumada, Jr "A standard model for foveal detection of spatial contrast," *Journal of Vision,* vol. 5, pp. 717-740, 2005.

[3] A. B. Watson, "A standard model for foveal detection of spatial contrast," *Journal of Vision,* vol. 5, pp. 714-740, 2005.

[4] X. Zhang and B. A. Wandell, "A spatial extension to CIELAB for digital color image reproduction," in *Society for Information Display Symposium Technical Digest*. vol. 27, ed, 1996, pp. 731-734.

[5] G. S. Brindley, *Physiology of the retina and the visual pathway*: Edward Arnold Ltd, London, 1960.

[6] D. Y. Teller and E. N. J. Pugh, "Linking propositions in color vision," in *Colour Vision: Physiology and Psychophysics*, J. D. Mollon and L. T. Sharpe, Eds., ed London: Academic Press, 1983.

[7] D. M. Green and J. A. Swets, *Signal Detection Theory and PsychophysicsNew York*. New York: wiley, 1966.

[8] W. S. Geisler, "Ideal observer analysis," in *Visual Neurosciences*, L. M. Chalupa and J. S. Wener, Eds., ed Cambridge MA: MIT Press 2003.

[9] W. S. Geisler, "Contributions to ideal observer theory to vision research," *Vision Research,* vol. 51, pp. 771-781, 2011.

[10] B. Wandell, D. H. Brainard, J. Winawer, J. E. Farrell, and H. Jiang. *ISETBIO: Tools for modeling image systems engineering in the human visual system front end*. Available: https://github.com/isetbio/isetbio

[11] J. Farrell, F. Xiao, P. Catrysse, and B. Wandell, "A simulation tool for evaluating digital camera image quality," *Proceedings of the SPIE,* vol. 5294, pp. 124-131, 2004.

[12] J. E. Farrell, P. B. Catrysse, and B. A. Wandell, "Digital camera simulation," *Journal of Applied Optics,* vol. 51, 2012.

[13] D. Brainard, H. Hofer, and B. Wandell. *Wavefront Optics: Matlab toolbox for analyzing wavefront optics data; especially human adaptive optics measurements*. Available: https://github.com/isetbio/WavefrontOptics

[14] J. Farrell, G. Ng, X. Ding, K. Larson, and B. Wandell, "A Display Simulation Toolbox for Image Quality Evaluation," " *IEEE/OSA Journal of Display Technology,* vol. 4, pp. 262-270, 2008.

[15] L. Thibos, "Retinal image quality for virtual eyes generated by a statistical model of ocular wavefront aberration," *Ophthalmic And Physiological Optics,* vol. 29, pp. 288-291, 2009.

[16] . *Color and Vision Research Laboratory*. Available: http://www.cvrl.org/

[17] D. H. Brainard, D. Pelli, and M. Kleiner. *Psychtoolbox Wiki: Home Page*. Available: http://psychtoolbox.org/HomePage

[18] W. Hammond B. R., B. R., Snodderly, D. M., , "Cigarette smoking and Retinal Carotenoids: Implications for Age-Related Macular Degeneration," *Vision Research,* vol. 36, pp. 3003-3009 1996.

[19] C.-C. Chang and C.-J. Lin, "LIBSVM : a library for support vector machines," *ACM Transactions on Intelligent Systems and Technology,* vol. 2, 2011.

[20] J. Burge. *Accuracy Maximization Analysis (AMA): Matlab and C code implementation* Available: http://jburge.cps.utexas.edu/research/Code.html

[21] J. Burge and W. S. Geisler, "Optimal defocus estimation in individual natural images.," *Proceedings of the National Academy of Sciences,* vol. 108, pp. 16849-16854, 2011.

[22] G. W., J. Najemnik , and A. D. Ing, "Optimal stimulus encoders for natural tasks. Journal of Vision, 9(13):17, 1-16 (2009)," vol. 9, pp. 1-16, 2009.

[23] O. Schade, "Optical and photoelectric analog of the eye," *Journal of the Optical Society of America,* vol. 46, pp. 721-739, 1956.

[24] B. A. Wandell, *Foundations of Vision*: Sinauer Associates, Inc., 1995.

[25] K. L. Gunther and K. R. Dobkins, "Individual differences in chromatic (red/green) contrast sensitivity are constrained by the relative number of L- versus M-cones in the eye," *Vision Research,* vol. 42, pp. 1367-1378, 2002.