Visibility of Uncorrelated Image Noise

Jiajing Xu[a], Reno Bowen[b], Jing Wang[c], and Joyce Farrell[a]
[a]Dept. of Electrical Engineering, Stanford University, Stanford, CA. 94305 U.S.A.
[b]Dept. of Psychology, Stanford University, Stanford, CA. 94305 U.S.A.
[c]Logitech Inc., 6505 Kaiser Drive, Fremont, CA. 94555 U.S.A.

ABSTRACT

In this study, we evaluated the effect that pixel size has upon people's preferences for images. We used multispectral images of faces as the scene data and simulated the response from sensors with different pixel size while the other sensor parameters were kept constant. Subjects were asked to choose between pairs of images; we found that preference judgments were primarily influenced by the visibility of uncorrelated noise in the images. We used the S-CIELAB metric (∆E) to predict the visibility of the uncorrelated image noise. The S-CIELAB difference between a test image and an ideal reference image was monotonically related to the preference score.

Keywords: sensor design, noise measurement, image quality metrics, S-CIELAB

1. INTRODUCTION

Camera manufacturers, vendors and customers use a variety of different metrics to quantify the image quality of a captured image. Common test targets, such as the Macbeth Color Checker (MCC) and the ISO12233 slanted edge, have played a useful role in comparing image quality across different cameras. For example, a common method for characterizing visible noise is to measure the variance in the image pixel values for one of the gray patches in the Macbeth Color Checker (MCC). And a common method for characterizing spatial resolution is to calculate the amplitude response to the different spatial frequencies present in a slanted bar [1, 2]. A limitation of these methods is that they do not account for the effects that display properties, viewing conditions and human visual sensitivity have upon the visibility of noise.

S-CIELAB is a spatial extension of the CIELAB metric that takes into account viewing distance, display properties and the spatial-chromatic sensitivities of the human observer [3, 4]. Previous studies have shown that S-CIELAB can predict the visibility of uncorrelated image noise [5, 6]. The main goal of the studies we report here is to determine whether we can use the S-CIELAB spatial-color difference metric to predict the effect that uncorrelated image noise has on subjective preference judgments.

A secondary goal of this study is to evaluate the effect that sensor pixel size has on preference judgments. We used the Image Systems Evaluation Toolbox (ISET) [7, 8] to simulate sensors with different pixel size. We kept other sensor parameters constant (such as the die size, fill factor, quantum efficiency, voltage swing, read noise, dark voltage, dark signal non-uniformity (DSNU) and photo-response non-uniformity (PRNU)). Although these parameters affect image quality, they were not varied in this experiment in order isolate the effects of pixel size. We used a diffraction-limited shift-invariant model for the optics and kept optical parameters, such as f# and focal length, constant.

In previous papers, we described how to use ISET to model and simulate the complete image processing pipeline of a digital camera, beginning with a radiometric description of the scene captured by the camera and ending with a radiometric description of the final image rendered on an LCD display [7, 8]. The ISET models based on a modest set of sensor measurements can predict the performance of real digital cameras [9].

In this study we used multispectral images of faces as the scene data and ISET to create and process sensor images [8]. Experiments were controlled using the Psych Toolbox [10], and the Display Simulation Toolbox [11] was used to predict

the displayed radiance of the images as they are rendered on a calibrated LCD in our laboratory. Finally, the S-CIELAB color difference metric was used to predict the visibility of uncorrelated image noise [12].

2. METHOD

2.1. Stimuli

The experimental images that subjects viewed were rendered by simulating the optics, sensor and image processing as described below.

Scene. We began with radiance images of two human faces; an Asian-American woman and an African-American male (see Figure 1). Each scene was rendered at three different mean luminance levels: 5, 10 and 15 cd/m^2.

Figure 1. RGB renderings of the two radiance images used as stimuli in the experiment

Table 1 lists the sensor parameters used in this study. Pixel size varies between 1.2 and 2.8 microns. We assume that the die size of the sensor array is the same for all sensors. Consequently, digital cameras with smaller pixels will have more pixels in the sensor array than cameras with larger pixels.

Optics. The scene data were optically blurred using wavelength dependent point spread functions derived from a diffraction-limited shift-invariant model. The lens f# and focal length were 2.0 mm and 3.7 mm, respectively.

Sensor. The well capacity was modeled as a function of the area of the pixel. Specifically well capacity is C*pixel_area, where C is derived from the manufacture-specified well capacity for a sensor with 2.8 micron pixels. According to the manufacturer, this sensor has a full well capacity of 17500 electrons. Thus, for this sensor, C is 2232 e-/um^2 .

The voltage swing for all sensors was assumed to be 1.2 volts and conversion gain was calculated for each sensor by dividing voltage swing by the well capacity. Read noise was held constant across all sensors at 10 electrons. Other sources of sensor noise (dark voltage, DSNU, and PRNU) are set to 0. Consequently, the main source of noise at low light levels will be read noise and the main source of noise at high light levels will be photon (shot) noise. The sensor conversion gain parameter converts the units from electrons to volts.

We assumed that each pixel had a microlens that can focus 90% of the light on the photosensitive region of a pixel (fill factor is 90%). The exposure duration was set to 15 ms. For this exposure duration, the highest voltage value across all sensor images is approximately half of the well capacity. We applied an analog gain to equate the sensor voltages so that

the intensity histograms are similar for all experimental images. This insures that the tone mapping is consistent for all sensor images at different scene luminance levels, and subjects' preferences are not determined by differences in tone mapping.

Table 1: Independent variables in the sensor simulations

Pixel size (microns)	2.8	2.2	1.7	1.4	1.2
Well capacity(e)	17500	10804	6451	4375	3214
Conversion gain	73	118	198	262	398
Read noise (mV)	0.731	1.185	1.984	2.926	3.982
Read noise (e-)	10	10	10	10	10

Image processing. The image processing pipeline consists of (a) bilinear demosaicking, performed separately in each color channel (b) color correction, and (c) image resizing (blurring and decimating). Uncorrelated image noise remains uncorrelated after this demosaicking method [13]. We color-balanced the sensor images using a color correction matrix (CCM) selected to optimized the rendering of a Macbeth color chart under the illuminant used to acquire the original scene [14].

Because of the difference in pixel size, the demosaicked sensor images have different sizes. The images were resized (Matlab resize function) to be 512x512. This function uses a blur and decimate algorithm to resize the images. The simulations represent a typical scenario in which customers view images captured by different cameras, but the images are rendered onto the same display with a similar intensity histogram.

2.2. Procedure

Four subjects completed a total of 1680 preference judgments, grouped into 8 blocks of 210 trials. In each block of trials subjects selected between pairs of images of the same scene (female Asian-American or male African-American face) with different mean luminance and a range of pixel sizes. Subjects viewed the images pairs on a Dell Model 1905P LCD. A chin rest kept the viewing distance constant (0.38 m).

3. Results

3.1. The effect of pixel size and scene luminance

An image preference score was calculated for each subject and stimulus condition. The score is the number of times an image was preferred over all the other images. Because there was very little difference between preference scores for the four different subjects and the two different scenes (Asian-American woman and African-American male), preference scores were averaged across subjects and scenes.

The mean and standard deviation of image preference scores vary as a function of pixel size with scene luminance (Figure 2). At low scene luminance levels (5 cd/m^2) preference scores increase with pixel size. As scene luminance increases, the dependence on pixel size is less pronounced, with preferences remaining nearly constant for pixel sizes between 1.8 and 2.8 microns.
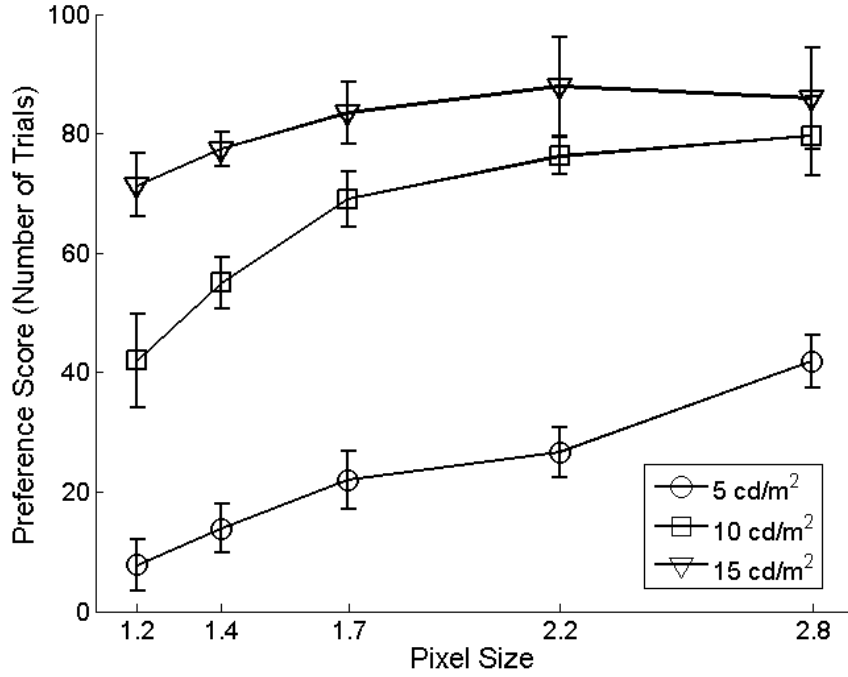
Figure 2. Preference scores plotted as a function of pixel size, with scene luminance as the parameter. The error bars are the standard error of the mean of the preference scores across the four observers.

The relationship between pixel size and scene luminance can be explained by the effect of read noise on the sensor signal-to-noise ratio (SNR) at low light levels. Smaller pixels capture fewer photons and consequently generate fewer electrons. The ratio of electrons generated by the signal and electrons generated by read noise (SNR) decreases with pixel size. The loss of SNR with pixel size is mitigated by downsizing the images. Images with more pixels (smaller pixel size) were resized to have the same size as the image with the fewest pixels. Downsizing compensates, in part, for the decrease in SNR. However, the preference scores show that downsizing does not completely offset the loss of SNR for small pixel sizes.

Subjects reported that their judgments were determined by the amount of visible noise each image. In the next section of this paper, we evaluate two metrics for quantifying the visibility of noise.

3.2. Noise Visibility Metrics

A common method for characterizing visible noise is to measure the variance in the camera RGB values for one of the gray patches in the MCC. These values are then used to calculate a measure that we refer to as the MCC_SNR. We used the ISET sensor simulation to predict the camera RGB values for the MCC target. The simulation parameters were identical to those used to generate the image for the visual preference task. The mean and variance of the RGB values for the 50% patch in the MCC were used to calculate the MCC_SNR

$$MCC\_SNR = 20 \cdot \log\left(\frac{Mean\_Pixel\_Value}{Std\_Deviation}\right)$$
(1)

Subjects' preference scores are not well predicted by the MCC SNR metric (Figure 3). Specifically, for images with the same MCC_SNR, preference scores can vary substantially.
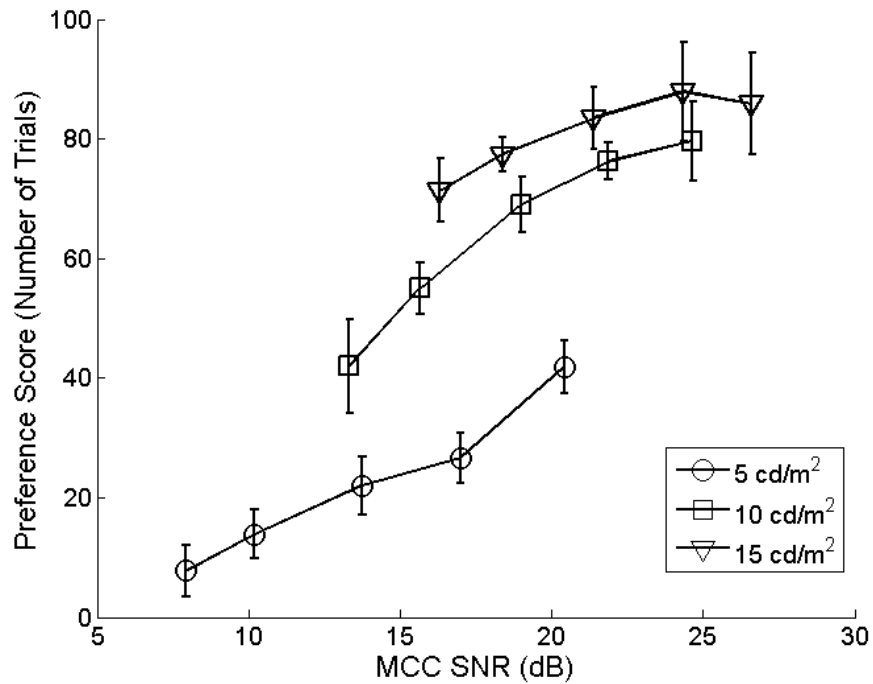

Figure 3. Mean and standard deviation in preference scores averaged across 4 subjects and plotted as a function of MCC SNR.

3.3. S-CIELAB predictions

S-CIELAB can be used as a reference metric. One can calculate the difference between a test image and an "ideal" reference image. We created an ideal reference image by capturing the scene data with a camera that required no downsizing and had the highest SNR (2.8 micron pixels, no noise sources and a mean scene luminance of 15 cd/m^2).

Figure 4 illustrates how the S-CIELAB difference between an image and the ideal reference image is calculated. The spatial properties of display pixels, the spectral power distribution of the display color primaries and the display gamma are used to calculate the displayed radiance of the test and ideal images as rendered on the target display [11]. The S-CIELAB difference (ΔE) between the two radiance images generates a spatial error (ΔE) map. The mean S-CIELAB difference (ΔE) (averaged across the error map) is a measure of the visibility of the difference between the test image and the ideal noise-free reference image.
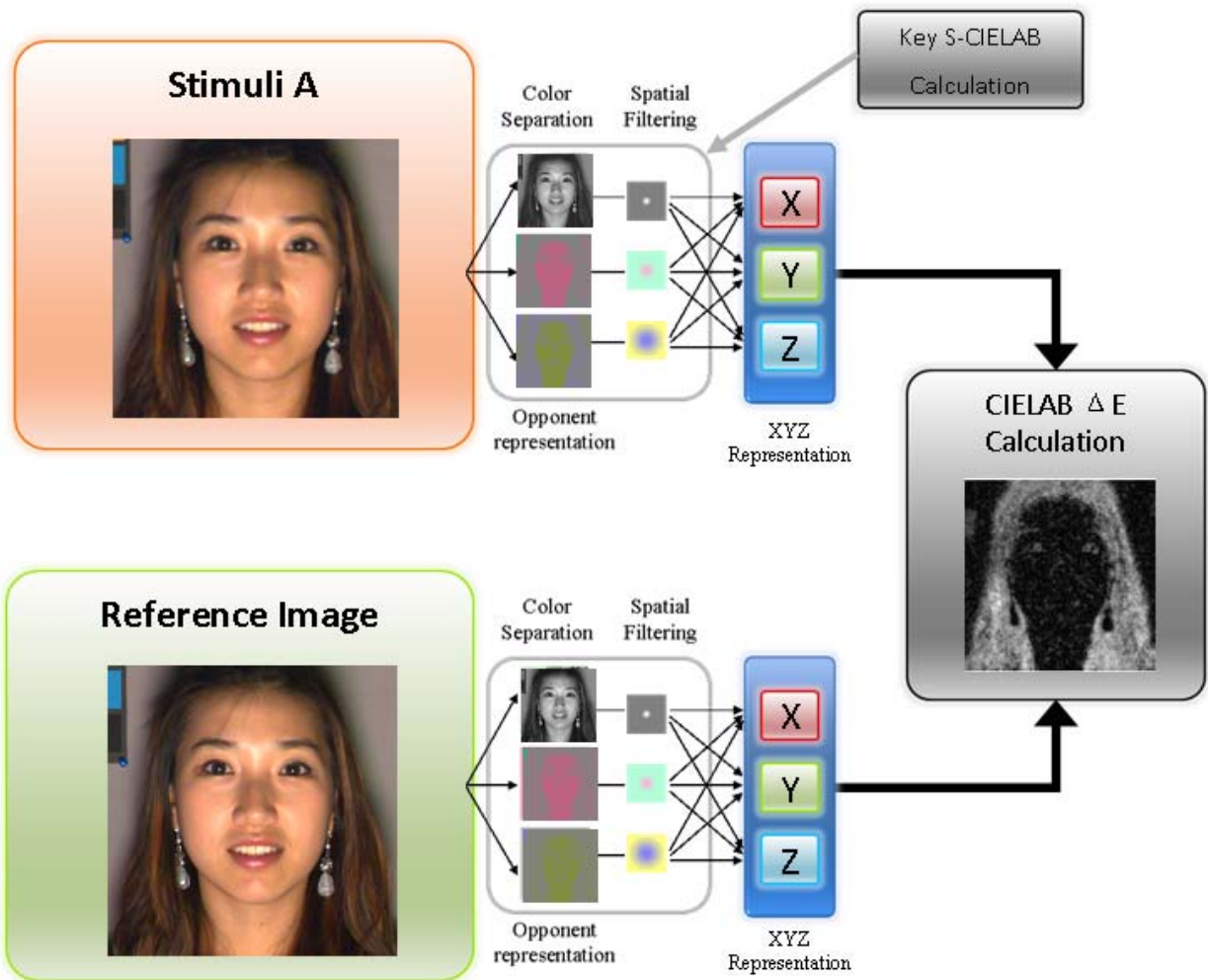
Figure 4. The S-CIELAB difference between two radiance images. The spatial error map is calculated by separating each image into separable luminance and chrominance components [15] and then blurring each component with a different contrast sensitivity function. After this spatial filtering process, the three separable image components are converted into the XYZ and Lab color representations, respectively. The S-CIELAB difference between two images is the Euclidean distance between two images in the Lab color space [3]. A summary measure is the mean of these difference values

There is a monotonically decreasing relationship between SCIELAB error (ΔE) and preference scores (Figure 5). The visibility of noise increases with the S-CIELAB difference (ΔE) between an ideal image and test image. This measure of noise visibility is monotonically related to the image preference score.
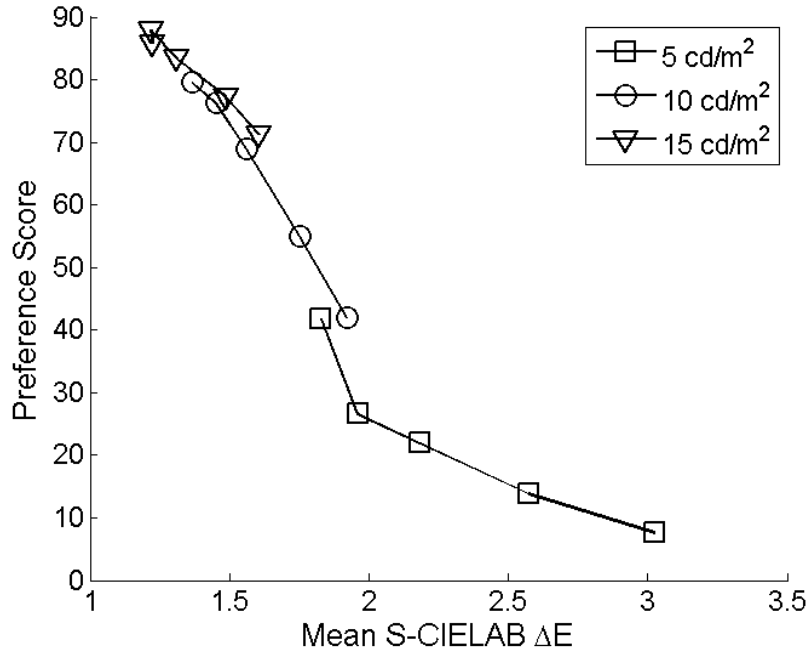
Figure 5. The mean S-CIELAB error is monotonically related to the image preference score. The data points show the mean preference scores, averaged across 4 subjects, as a function of S-CIELAB ΔE. The error bars are the standard error of the mean of the preference scores across the four observers.

4. DISCUSSION

Visual psychophysical experiments are an excellent way to quantify perceived image quality; but they are labor-intensive and time-consuming. It is desirable, therefore, to have a computational method to predict image quality, without requiring visual psychophysical experiments. The S-CIELAB metric predicted the preference judgments of the four individuals that served as subjects in our experiments. Specifically, the mean S-CIELAB difference (ΔE) between a test image and an ideal noise-free reference image is monotonically related to the preference score derived from subjects' judgments. Furthermore, the mean S-CIELAB difference (ΔE) metric is more closely correlated to subjects' preference scores than the MCC SNR metric. The S-CIELAB metric takes into account the spatial and spectral properties in the images that subjects view, the display upon which the images are rendered, the viewing distance and the spatial-chromatic sensitivities of the human observer.

It is too soon to generalize the results in this study much beyond the scene, optical, sensor and image processing parameters used in the ISET simulations and experiments. There are a variety of types of image quality measurements. In this experiment subjects reported that their judgments were influenced mainly by the amount of visible noise in the images. Other differences, such as sharpness or color tone, are not likely to be captured by this metric.

For these types of noise-limited measurements, ISET and S-CIELAB are useful tools that can predict how certain properties of a scene and sensor influence subject's preference judgments. As an example, Figure 6 shows the S-CIELAB predictions as a function of pixel size and scene luminance. The graph on the left shows the predictions for a sensor with read noise of 10 electrons. This graph predicts that the visibility of noise decreases with increasing pixel size and scene luminance, as we found in our visual psychophysical experiments. The graph on the right shows the S-CIELAB predictions for a sensor that has no read noise. In this case, pixel size does not influence the visibility of noise. The increase in SNR with downsizing offsets the decrease in SNR due to photon noise such that there is no effect of pixel size.
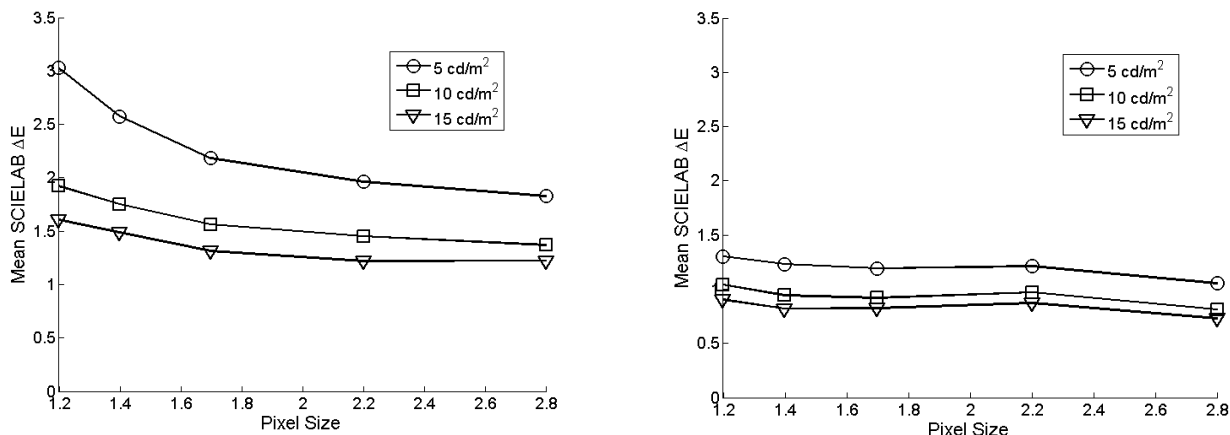
Figure 6. The effect of read noise on predicted preference for different pixel sizes. S-CIELAB ΔE values are shown as a function of pixel size and scene luminance. The two graphs are calculated assuming different levels of read noise. On the left read noise was set to 10 electrons; on the right graph read noise was set to zero

5. Acknowledgments

REFERENCES

[1] Standardization, I.O.f., *Photography-Electronic Still Picture Cameras-Resolution Measurements*. 1999: New York, New York, USA.

[2] Williams, D., *Benchmarking of the ISO 12233 Slanted-Edge Spatial Frequency Response (SFR) Plug-in*, in *Proceedings of the 51st IS&T PICS Conference*. 1988. p. 133-136.

[3] Zhang, X. and B. Wandell, *A spatial extension to CIELAB for digital color image reproduction*, in *Society for Information Display Symposium Technical Digest*. 1996. p. 731-734.

[4] Zhang, X., J. Farrell, and B. Wandell, *Applications of S-CIELAB: A spatial extension to CIELAB*, in *Proceedings of the {IS\&T/SPIE} 9th Annual Symposium on Electronic Imaging*. 1997.

[5] Zhang, X., D.A. Silverstein, J. Farrell and B. Wandell., *Color image quality metric S-CIELAB and its application on halftone texture visibility*, in *{COMPCON97} Digest of Papers*. 1997. p. 44-48.

[6] Zhang, X. and B. Wandell, *Color image fidelity metrics evaluated using image distortion maps*. Signal Processing, 1998. **70**(3): p. 201-214.

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

[8] Farrell, J., M. Parmar, P. Catrysse and B. Wandell, *Digital Camera Simulation*, in *Handbook of Digital Imaging*, M. Kriss, Editor. in press, Wiley.

[9] Farrell, J., M. Okincha, and M. Parmar, *Sensor calibration and simulations*. Proceedings of the SPIE, 2008. **6817**.

[10] Brainard, D.H., *The psychophysics toolbox.* Spatial Vision, 1997. **10**: p. 433-436.

[11] Farrell, J., G. Ng, X. Ding, K. Larson and B. Wandell, *A Display Simulation Toolbox for Image Quality Evaluation,"* IEEE/OSA Journal of Display Technology, 2008. **4**(2): p. 262-270.

[12] Zhang, X. *S-CIELAB: A spatial extension to the CIE L*a*b* DeltaE Color Difference Metric*. 1998 [cited; Available from: http://white.stanford.edu/~brian/scielab/scielab.html.

[13] Park, S.H., H.S. Kim, S. Lansel, M. Parmar and B. Wandell, *A case for denoising before demosaicking color filter array data*, in *43rd Annual Asilomar Conference on Signals, Systems, and Computers*. 2009: Monterey, California.

[14] ImagEval. *ISET - Selecting a color conversion matrix*. 2009 [cited; Available from: http://www.imageval.com/public/Products/ISET/ApplicationNotes/ColorCorrectionMatrix.pdf.

[15] Poirson, A. and B. Wandell, *Pattern-color separable pathways predict sensitivity to simple colored patterns*. Vision Research, 1996. **36**(4): p. 515-526.