

# Measuring Images: Differences, Quality, and Appearance

Garrett M. Johnson<sup>□</sup> and Mark D. Fairchild

Munsell Color Science Laboratory, Chester F. Carlson Center for Imaging Science,  
Rochester Institute of Technology, Rochester, NY, USA 14623-5604

## ABSTRACT

One goal of image quality modeling is to predict human judgments of quality between image pairs, without needing knowledge of the image origins. This concept can be thought of as device-independent image quality modeling. The first step towards this goal is the creation of a model capable of predicting perceived magnitude differences between image pairs. A modular color image difference framework has recently been introduced with this goal in mind. This framework extends traditional CIE color difference formulae to include modules of spatial vision and adaptation, sharpness detection, contrast detection, and spatial localization. The output of the image difference framework is an error map, which corresponds to spatially localized color differences. This paper reviews the modular framework, and introduces several new techniques for reducing the multi-dimensional error map into a single metric. In addition to predicting overall image differences, the strength of the modular framework is its ability to predict the distinct mechanisms that cause the differences. These mechanisms can be thought of as attributes of image appearance. We examine the individual mechanisms of image appearance, such as local contrast, and compare them with overall perceived differences. Through this process, it is possible to determine the perceptual weights of multi-dimensional image differences. This represents the first stage in the development of an image appearance model designed for image difference and image quality modeling.

Keywords: Color image differences, image appearance, image quality, vision modeling

## 1. INTRODUCTION

Techniques for image quality modeling can be described as falling into two categories: Device-dependent and Device-independent. Device-dependent models relate imaging system parameters such as dots-per-inch (DPI), contrast, and gamut volume to human perceptions. These perceptions might be individual attributes such as sharpness and graininess, or overall image quality, and are typically related through psychophysical experimentation. These techniques are considered device-dependent, as they are only valid for a given imaging system. The relationship must be recalculated if the system is changed. Keelen has presented an overview of this type of image quality modeling.<sup>1</sup>

Device-independent image quality techniques attempt to utilize information contained in the images themselves, rather than knowledge of the imaging system. Often this is accomplished through modeling of the human visual system. These models can further be categorized as predictors of thresholds or magnitudes. This research focuses on the design and formulation of a framework for device independent image quality research. The pedigree of this framework stems from the CIE color difference equations, combined with spatial models of vision such as the S-CIELAB spatial pre-processing.<sup>2</sup>

This paper describes the evolution of a color image difference framework into a model of image appearance. An image appearance model can be thought of a color appearance model for complex spatial stimuli. This allows for the prediction of appearance attributes such as lightness, chroma, and hue, as well as image attributes such as sharpness, contrast, and graininess. The prediction of these attributes can then be used to formulate a device-independent metric for overall image quality.

### 1.1 Color Difference Equations

Color difference research has culminated with the recently published CIEDE2000 color difference formula.<sup>3</sup> A color difference equation allows for the mapping of physically measured stimuli into perceived differences. At the heart of the color difference equations lies a uniform color space. The CIE initially recommended two such color spaces, in 1976,

---

<sup>□</sup> garrett@cis.rit.edu, mdf@cis.rit.edu, www.cis.rit.edu/mcsl

CIELAB and CIELUV. Both spaces were initially described as interim color spaces, with the knowledge that they were far from complete. Over 25 years later these spaces are still the CIE recommendations, although CIELUV has fallen out of favor.

With a truly uniform color space, color differences can then be taken to be a simple measure of distance between two colors in the space, such as CIE  $\Delta E^*_{ab}$ . The CIE recognized the non-uniformity of the CIELAB color space, and formulated more advanced color difference equations such as CIE  $\Delta E_{94}$  and CIEDE2000. These more complicated equations are very capable of predicting perceived color differences of simple color patches.

### 1.2 Image Differences

The CIE color difference formula were developed using simple color patches in controlled viewing conditions. There is no reason to believe that they are adequate for predicting color difference for spatially complex image stimuli. The S-CIELAB model was designed as a spatial pre-processor to the standard CIE color difference equations, to account for complex color stimuli such as half-tone patterns.<sup>2</sup> The spatial pre-processing uses separable convolution kernels to approximate the contrast sensitivity functions (CSF) of the human visual system. The CSF serves to remove information that is imperceptible to the visual system. For instance, when viewing halftone dots at a certain distance the dots tend to blur, and integrate into a single color. A pixel-by-pixel color difference calculation between a continuous image and a halftone image would result in very large errors, while the perceived difference might in fact be small. The spatial pre-processing would blur the halftone image so that it more closely resembles the continuous tone image.

S-CIELAB represents the first incarnation of an image difference model based upon the CIELAB color space and color difference equations. Recently this model has been refined and extended into a modular framework for image color difference calculations.<sup>4</sup> This framework refines the CSF equations from the S-CIELAB models, and adds modules for spatial frequency adaptation, spatial localization, and local and global contrast detection.<sup>5</sup> This framework will be discussed in more detail below.

### 1.3 Color & Image Appearance models

A model capable of predicting perceived color difference between complex image stimuli is a useful tool, but has some limitations. Just as a color appearance model is necessary to fully describe the appearance of color stimuli, an image appearance model is necessary to describe spatially complex color stimuli. Color appearance models allow for the description of attributes such as lightness, brightness, colorfulness, chroma, and hue. Image appearance models extend upon this to also predict such attributes as sharpness, graininess, contrast, and resolution.

A uniform color space also lies in the heart of the of an image appearance model. The modular image difference framework allows for great flexibility in the choice of color spaces. Examples are the CIELAB color space, similar to S-CIELAB, the CIECAM02 color appearance model, or the IPT color space.<sup>6,7</sup>

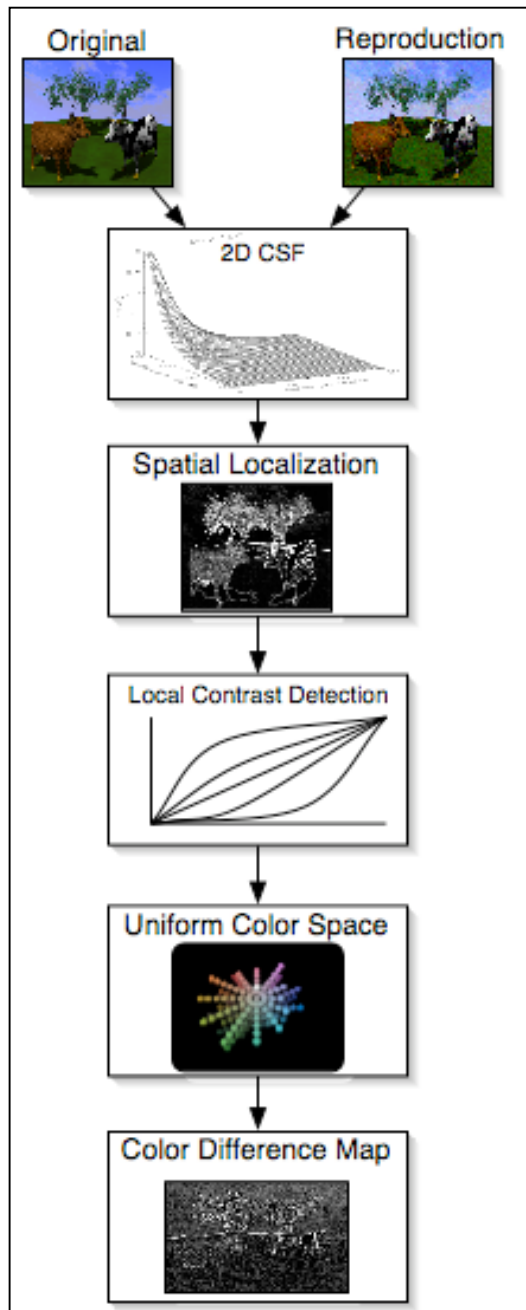
### 1.4 Image Quality

Models of image appearance can be used to formulate multi-dimensional models of image quality. For example it is possible to take weighted sums of various appearance attributes to determine a metric of overall image quality, as described by Keelen<sup>1</sup> and Engledrum.<sup>13</sup> Essentially these models can augment or replace human observations to weight image attributes with overall appearances of quality. For instance a model of quality might involve weighted sums of tonal balance, contrast, and sharpness. A first step towards this type of model is illustrated in more detail below.

## 2. MODULAR IMAGE DIFFERENCE MODEL

A framework for a color image difference metric has recently been described.<sup>4,5</sup> This framework was designed to be modular in nature, to allow for flexibility and adaptation. The framework itself is based upon the S-CIELAB spatial extension to the CIELAB color space. S-CIELAB merges traditional color difference equations with spatial properties of the human visual system. This was accomplished as a spatial filtering pre-processing, before a pixel-by-pixel color difference calculation.<sup>2</sup>

The modular framework further extends this idea by adding several pre-processing steps, in addition to the spatial filtering. These pre-processing steps are contained in independent modules, so they can be tested and refined. Several



**Figure 1. Flowchart of Modular Image Difference Metric.**

modules have been described, and include spatial filtering, adaptation, and localization, as well as local and global contrast detection. Figure 1 shows a general flowchart with several distinct modules. These modules are described briefly below.

#### *Spatial Filtering*

The behavior of the human visual system in regards to spatially complex stimuli has been well studied over the years.<sup>6</sup> The contrast sensitivity function describes this behavior in relation to spatial frequency. Essentially the CSF is described in a post-retinal opponent color space, with a band-pass nature for the luminance channel and low-pass nature for the chrominance channels. S-CIELAB uses separable convolution kernels to approximate the CSF, and modulate image details that are imperceptible. More complicated contrast sensitivity functions that include both modulation and frequency enhancement were discussed in detail by Johnson and Fairchild.<sup>5</sup>

#### *Spatial Frequency Adaptation*

The contrast sensitivity function in this framework serves to modulate spatial frequencies that are not perceptible, and enhance certain frequencies that are most perceptible. Generally CSFs are measured using simple grating stimuli with care taken to avoid spatial frequency adaptation. Spatial frequency adaptation essentially decreases sensitivity to certain frequencies based upon information present in the visual field.

Since spatial frequency adaptation cannot be eliminated in real world viewing conditions, several models of spatial frequency adaptation have been described.<sup>5</sup> These models alter the nature of the CSF based upon either assumptions of the viewing conditions, or based upon the information contained in the images themselves.

#### *Spatial Localization*

The band-pass and low-pass contrast sensitivity serve to modulate high-frequency information, including high-frequency edges. The human visual system is generally acknowledged to be very adept at detecting edges. To accommodate this behavior, a module of spatial localization has been developed. This module can be as simple as an image processing edge-enhancing kernel, although that kernel must change as a function of viewing distance. Alternatively, the CSF can be modified to boost certain high frequency information.

#### *Local Contrast Detection*

This module serves to detect local and global contrast changes between images. It is based upon the nonlinear mask based local contrast enhancement described by Moroney.<sup>9</sup> Essentially a low-pass image mask is used to generate a series of tone-reproduction curves. These curves are based upon the global contrast of the image, as well as the relationship between a single pixel and its local neighborhood.

#### *Color Difference Map*

The output of the modular framework is a map of color differences, corresponding to the perceived magnitude of error at each pixel location. This map can be very useful for determining specific causes of error, or for detecting systematic

errors in a color imaging system. Often times it is useful to reduce the error map into a more manageable dataset. This can be accomplished using image statistics, so long as care is taken. Such statistics can be image mean, max, median, or standard deviation. Different statistics might be more valuable than others depending on the application, as perhaps the mean error better describes overall difference, while the max might better describe threshold differences.

### 3. PSYCHOPHYSICAL DATASETS

Two psychophysical experiments were performed to aid in the evaluation and development of the image difference framework. The first experiment, described by Johnson and Fairchild, tested the effect of resolution, contrast, noise, and spatial filtering on the perception of sharpness.<sup>10</sup> The second experiment, described by Calabria and Fairchild, tested the effect of lightness, chroma, and sharpness manipulations on perceived contrast.<sup>11</sup> These experiments are reviewed below.

#### Sharpness Experiment

A large-scale experiment testing simultaneous manipulations on resolution, contrast, additive noise, and spatial filtering was performed. Overall there were 3 levels of resolution corresponding to 300, 150, and 100 pixels-per-inch, 3 levels of contrast enhancement, 4 levels of additive noise, and 2 levels of spatial filtering. This corresponded to 72 manipulations. These manipulations were performed on 4 image scenes, shown below. Overall there were 10,224 image pairs to be viewed. A total of 51 observers performed over 150,000 paired comparison evaluations.



Figure 2. Image Scenes From Sharpness Experiment

#### Contrast Experiment

Three separate experiments were performed to evaluate the effect of lightness, chroma, and sharpness on perceived contrast. The first experiment tested 20 manipulations on the CIELAB L\* channel, the second tested 7 manipulations on the CIELAB C\* channel, while the third tested 8 levels of spatial sharpening. For each experiment both preference and perceived contrast were judged using a paired comparison method. A total of 6 scenes were chosen for the experiments, which are shown below. There were approximately 30 observers for each of the three experiments.



Figure 3. Image Scenes From Contrast Experiment

Both the sharpness and contrast experiments were performed on a colorimetrically characterized Apple 22 inch Cinema display, which had an average error of less than 0.5 CIE  $\Delta E_{94}$ . The experimental analysis, performed using Thurstone's Law of Comparative Judgments, resulted in magnitude scales of sharpness and contrast.

#### 4. IMAGE DIFFERENCE PREDICTIONS

The contrast and sharpness scales from the psychophysical experiments can be used to test the image difference framework. Each of the experiments involves an "original" image, and a series of manipulations performed on the image. We can use the image difference framework to calculate this difference, and then compare the results against the experimental data. Since the image difference metric can only predict relationships between images, we must first normalize the magnitude experimental scales such that they represent scale differences between any image, and the original image. It is important to note that the image difference metric is only capable of predicting the magnitude of differences, and not the direction. That is to say, the model is unable to determine the appearance changes of the image. As such, two images might have the same measured difference from an original, although one is perceived to be sharper, while the other is perceived to be less sharp. An example of this relationship is shown below.

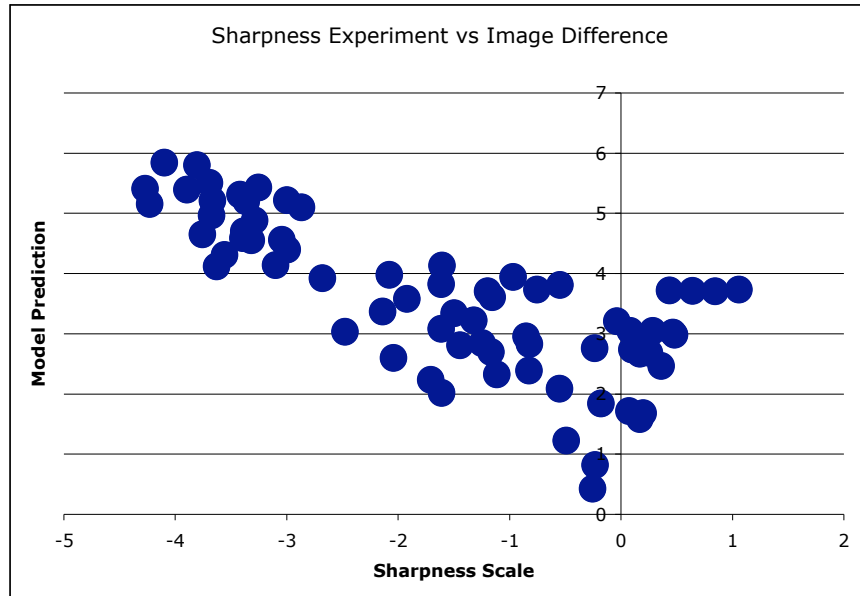
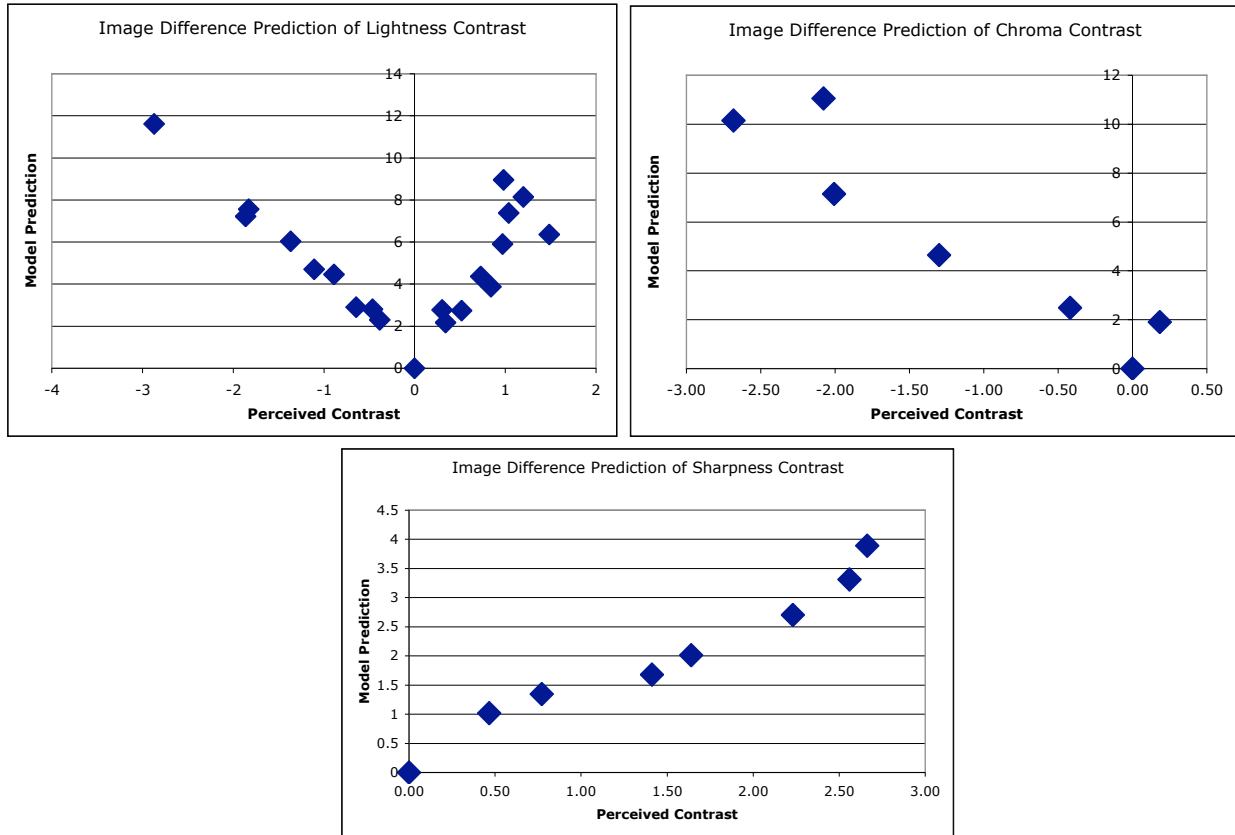


Figure 4. Model prediction of sharpness experiment data.

The above plot shows the prediction of the image difference model for the sharpness experiment. In order to relate the output error map against the magnitude experimental scale we must first reduce the dimensionality of the error. The model prediction shown in Figure 4 was calculated by taking the mean value of the error map, and is in units of CIEDE2000. It is important to note that the relative size of the data points plotted in Figure 4 are indicative of the 95% confidence interval of the experimental data.

There are several important aspects that are revealed in Figure 4. Due to the normalization of the magnitude scales to the original we have a distinct "V" shaped curve with the origin at zero. Any point to the left of the origin represents an image judged to be less sharp than the original, while any point to the right indicates an image judged to be sharper. Another important aspect is the spread of the prediction. The image difference model does an admirable job at predicting the general trend of the experiment, considering it is based entirely on the theory that color difference perceptions are highly correlated with sharpness. The multidimensional nature of the sharpness experiment also might have added noise to the overall dataset.

The contrast experiments were single dimensional, in that they varied only a single dimension and tested the perception of contrast based upon that change. An image difference metric should be better suited to predicting this type of dataset. Figure 5 shows the model predictions for the contrast experiment.



**Figure 5. Image difference model predictions of contrast experiments. Upper left is lightness manipulations, upper right is chroma manipulations, bottom is sharpness manipulations.**

The image difference model predicts the contrast experiment data very well, as can be seen by the nearly linear plots. Once again, the size of the data point is indicative of the experimental confidence interval. The lightness manipulations show a general “V” shaped trend, with a very tight grouping for the images judged to be of less contrast than the original. For the chroma manipulations there is a single outlying point, which corresponds to the image with only 20% chroma. This experimental evaluation of this image was very different than all the other images, as described by Calabria.<sup>11</sup> For the sharpness experiment all images were judged to have more contrast than the original, indicating a close relationship between sharpness and perceived contrast.

The above plots indicate the power of an image difference metric in predicting experimental data. The metric is especially strong at predicting the single dimensional dataset of the contrast experiment. The model struggles some with the multi-dimensional sharpness experiment, although it is capable of predicting the general trend well. The image difference is incapable of determining the direction of the differences. For that we need to move to a model capable of predicting appearance attributes.

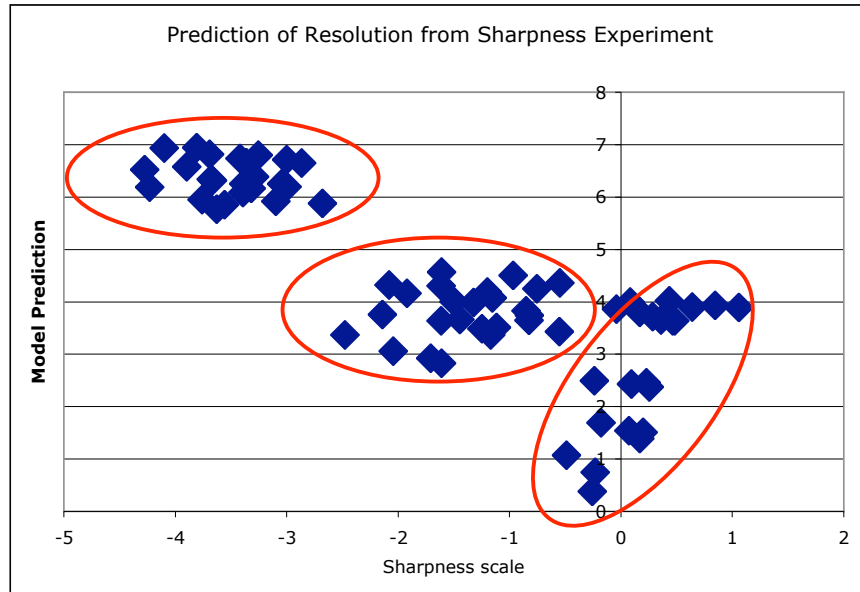
## 5. APPEARANCE ATTRIBUTE PREDICTIONS

One of the strengths of the modular image difference framework is the ability to pull out information from each module, without affecting any of the other calculations. This flexibility can be very valuable for determining causes of perceived difference, or for predicting attributes of image appearance. This section illustrates the strength of these techniques,

through the ability to pull-out many of the individual manipulations (resolution, contrast, and spatial filtering) from the sharpness experiment described above.

#### Resolution Prediction

It is well known that the luminance channel is much more sensitive to high frequency information than the chrominance channels. To predict resolution, we can look at the amount of error information in a luminance channel. This is easily accomplished by examining the standard deviation of the error in the CIELAB  $L^*$  channel, as the low frequency error will be small, while the higher frequency error will be very large. The plot of standard deviation of the  $L^*$  channel against the experimental sharpness scale is shown below.

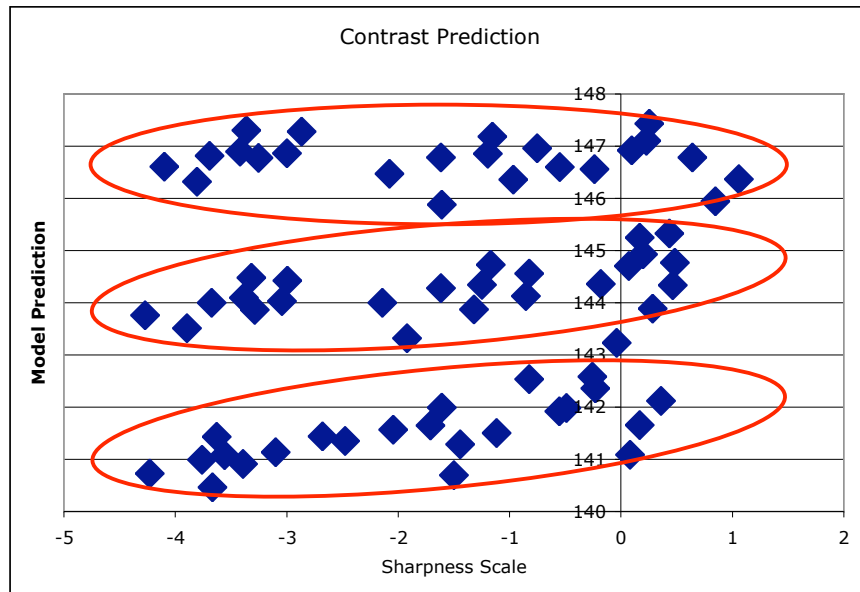


**Figure 6. Standard deviation of CIELAB  $L^*$  channel plotted against sharpness scale.**

There are three distinct groupings in Figure 6, which correspond the three levels of resolution. It should be noted that there is some noise in the higher resolution images, indicating that perhaps some of the other attributes are masking the perception of resolution. This noise might make it difficult for a model pull out the resolution information on its own, but this type of analysis is still a useful visual tool for a researcher. By examining a plot such as Figure 6 in addition to an image difference plot such as Figure 4, it should be possible to determine that resolution is the cause of the cause difference. This is the type of analysis that will be necessary to create more complicated multi-dimensional image quality models.

#### Contrast Prediction

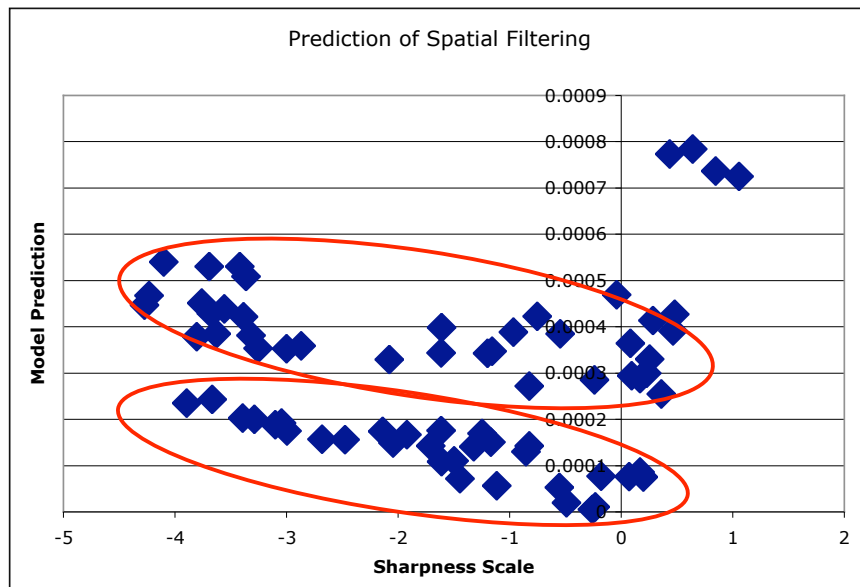
By examining the output of the mask-based contrast module it is possible to reveal the effect of contrast on the sharpness perception. The contrast module uses a low-pass mask to generate a series of tone curves based upon both global and local changes of contrast. The degree of the low-pass filter determines the local contrast neighborhood. Typically this is performed only on the luminance information, although a similar type metric could be used to determine changes in chroma contrast. To detect changes in contrast we can examine at the mean difference of the CIELAB  $L^*$  channel output from the contrast module, as shown in Figure 7. There are three distinct groups illustrated in Figure 7, corresponding to the three levels of contrast manipulation that were performed in the sharpness experiment. This type of analysis is easily able to separate the distinct levels with very little noise. It should be noted that this type of analysis is not attempting to predict the actual sharpness experiment, but is rather attempting to provide insight into the appearance attributes that caused the perceived differences.



**Figure 7. Mean of CIELAB  $L^*$  channel from contrast module**

*Spatial Filtering Prediction*

The spatial localization module is capable of predicting the effects of the spatial filtering. The localization module is essentially a band-pass filter centered at a specific frequency. This filter can be described as a Gaussian filter. Depending on the desired effect different Gaussian functions can be chosen. For example, a Gaussian of width 10 cycles-per-degree (cpd), centered at 30 cycles-per-degree can be used on the opponent luminance channel. The standard deviation of the luminance channel filtered by a Gaussian centered at 20 cpd, with a width of 5 cpd is shown below.



**Figure 8. Standard deviation of luminance channel filtered by spatial localization module.**

There are two distinct groups revealed in Figure 8, corresponding to the two levels of spatial filtering in the sharpness experiment. There are several outlying points that should lie with the upper group, indicating some noise. A simple

threshold on the standard deviation would be able to determine whether spatial filtering was applied to a given experimental image.

The spatial localization module can be used for pulling out the spatial filtering as shown above, and can also be useful for other predictions. By altering the height, width, and location of the Gaussian function it is possible to use the output of this module to predict the sharpness experimental scale on its own. This is illustrated in Figure 9, calculated with a Gaussian of width 10 cpd, centered at 20 cpd, with a height of 0.6. The mean of the luminance opponent channel filtered by this module is shown below. The general grouping of the model prediction using just the spatial localization module appears to be much tighter than the mean color difference grouping shown in Figure 4. This indicates that perhaps a color difference equation does not fully capture the perception of sharpness.

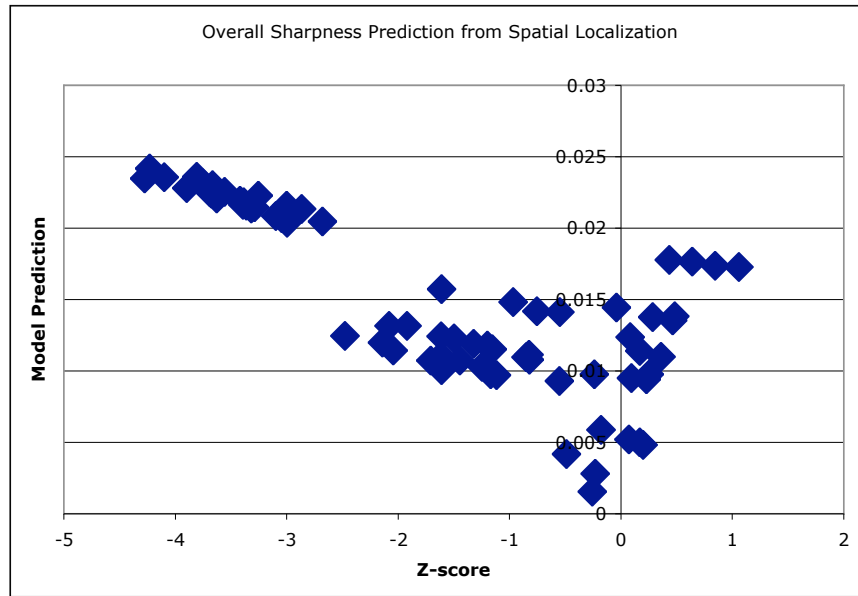
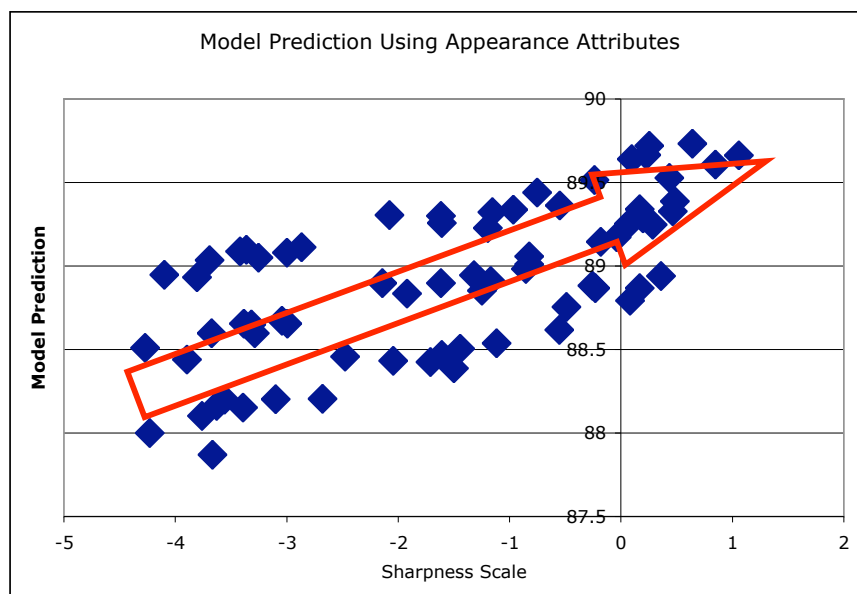


Figure 9. Mean of luminance channel filtered by a Gaussian of width 10cpd, centered at 20 cpd.

## 6. TOWARDS IMAGE APPEARANCE AND IMAGE QUALITY MODELS

The above section illustrates the goals of a model of image appearance capable of predicting both image differences, and appearance attributes such as sharpness, and contrast. Once such model has recently been described by Fairchild and Johnson.<sup>12,13</sup> This model extends the modular framework by replacing the color space selection with the appearance space IPT.

With an image appearance model in place, it should be possible to construct a multi-dimensional model of image quality using the techniques described by Keelan<sup>1</sup> and Engledrum.<sup>14</sup> These type of models weigh various appearance attributes together to form a single metric of quality. An example of this type of calculation is shown in Figure 10. This calculation is still, in essence, a mean color difference calculation but is enhanced by weighting the various appearance attributes from the independent modules. The resulting color difference calculation is no longer characterized by the “V” shape, but rather has a monotonic relationship with the sharpness experiment. This indicates that the model is capable of predicting both magnitude and direction of color differences. There is still a spread in the data, but the general trend indicated by the arrow is very evident. The three distinct lines are representative of the differences in contrast, indicating perhaps too much weight from the contrast module.



**Figure 10. Example of a Multi-dimensional Image Quality Model.**

## 7. CONCLUSIONS

We have described the evolution from models of image difference, through image appearance, and ultimately image quality. These models fuse spatial vision research with color difference and color appearance calculations. The results of model prediction from several psychophysical experiments were discussed.

## 8. REFERENCES

1. B.W. Keelan, *Handbook of Image Quality: Characterization and Prediction*, Marcel Dekker, New York, NY (2002).
2. X. M. Zhang and B. A. Wandell, A spatial extension to CIELAB for digital color image reproduction, *Proceedings of the SID Symposiums*, 731-734 (1996).
3. M. R. Luo, G. Cui, and B. Rigg, "The development of the CIE 2000 Colour Difference Formula, *Color Research and Applications*, **26**, 340-350 (2000).
4. G.M. Johnson and M.D. Fairchild, Darwinism of Color Image Difference Models, *Proc. of IS&T/SID 9<sup>th</sup> Color Imaging Conference*, 108-112 (2001).
5. G.M. Johnson and M.D. Fairchild, On Contrast Sensitivity in an Image Difference Model, *Proc. of IS&T PICS Conference*, 18-23 (2001).
6. B.A. Wandell, *Foundations of Vision*, Sinauer Associates Inc. Sunderland, MA (1995).
7. N. Moroney *et al*, "The CIECAM02 Color Appearance Model," *Proc. of IS&T/SID 10<sup>th</sup> Color Imaging Conference*, 23-27 (2002).
8. F. Ebner and M.D. Fairchild, Development and testing of a color space (IPT) with improved hue uniformity, *Proc of IS&T/SID 6<sup>th</sup> Color Imaging Conference*, 8-13 (1998).
9. 12 N. Moroney, Local Color Correction Using Non-Linear Masking, *Proc. of IS&T/SID 8th Color Imaging Conference*, 108-111 (2000).
10. G.M. Johnson and M.D. Fairchild, Sharpness Rules, *Proc of IS&T/SID 8<sup>th</sup> Color Imaging Conference*, 24-30 (2000).
11. A.J. Calabria and M.D. Fairchild, Compare and Contrast: Perceived Contrast of Color Images, *Proc. of IS&T/SID 10<sup>th</sup> Color Imaging Conference*, 17-22 (2002).
12. M.D. Fairchild and G.M. Johnson, Meet iCAM: An Image Appearance Model, *IS&T/SID 10<sup>th</sup> Color Imaging Conference*, 33-38 (2002).
13. M.D. Fairchild and G.M Johnson, Image Appearance Modeling, *Proc. SPIE/IS&T Electronic Imaging Conference*, in press, (2003).
14. P.G.Engledrum, "Extending Image Quality Models," *Proc IS&T PICS Conference*, 65-69 (2002).