Rochester Institute of Technology

RIT Scholar Works

Theses

5-1-2003

Measuring images: Differences, quality and appearance

Garrett M. Johnson

Follow this and additional works at: <https://scholarworks.rit.edu/theses>

Recommended Citation

Johnson, Garrett M., "Measuring images: Differences, quality and appearance" (2003). Thesis. Rochester Institute of Technology. Accessed from

This Dissertation is brought to you for free and open access by RIT Scholar Works. It has been accepted for inclusion in Theses by an authorized administrator of RIT Scholar Works. For more information, please contact ritscholarworks@rit.edu.

MEASURING IMAGES: DIFFERENCES, QUALITY AND APPEARANCE

Garrett M. Johnson

M.S. Color Science, Rochester Institute of Technology
(1998)

A dissertation submitted in partial fulfillment of the
requirements for the degree of Ph.D.
in the Chester F. Carlson Center for Imaging Science
of the College of Science
Rochester Institute of Technology

May 2003

Signature of the Author _____

Accepted By _____ *6/12/03*

Coordinator, Ph.D. Degree Program

Date

CHESTER F. CARLSON
CENTER FOR IMAGING SCIENCE
COLLEGE OF SCIENCE
ROCHESTER INSTITUTE OF TECHNOLOGY
ROCHESTER, NEW YORK

CERTIFICATE OF APPROVAL

Ph.D. DEGREE DISSERTATION

The Ph.D. Degree Dissertation of Garrett M. Johnson
has been examined and approved by the
dissertation committee as satisfactory for the
dissertation requirement for the
Ph.D. degree in Imaging Science

Prof. Ed Granger, Chair

Prof. Mark D. Fairchild, Thesis Advisor

Prof. Jeff Pelz

Prof. Jon Arney

Prof. Ethan Montag

THESIS RELEASE PERMISSION
ROCHESTER INSTITUTE OF TECHNOLOGY
COLLEGE OF SCIENCE
CHESTER F. CARLSON
CENTER FOR IMAGING SCIENCE

Title of Thesis: MEASURING IMAGES: DIFFERENCES, QUALITY, AND APPEARANCE

I, Garrett M. Johnson, hereby grant permission to the Wallace Memorial Library of R.I.T. to reproduce my thesis in whole or in part. Any reproduction will not be for commercial use or profit.

Signature: _____

Date: 4/3/03

MEASURING IMAGES: DIFFERENCES, QUALITY AND APPEARANCE

Garrett M. Johnson

M.S. Color Science

(1998)

A dissertation submitted in partial fulfillment of the
requirements for the degree of Ph.D.
in the Chester F. Carlson Center for Imaging Science
of the College of Science
Rochester Institute of Technology

May 2003

ABSTRACT

In order to predict the overall perception of image quality it is necessary to first understand and quantify the appearance of images. Just as color appearance modeling evolved from traditional colorimetry and color difference calculations, image appearance modeling evolves from color image difference calculations. A modular framework for the creation of a color image difference metric has been developed and tested using several psychophysical datasets. This framework is based upon traditional CIE color difference equations, and the S-CIELAB spatial extension to the CIELAB color space. The color image difference predictions have been shown to correlate well with experimental data. The color image difference framework was extended predict the overall appearance of images by replacing the CIELAB color space at the heart of the calculations with a color appearance space. An image appearance model maps the physics of complex image stimuli into human perceptions such as lightness, chroma, hue, contrast, sharpness, and graininess. A first generation image appearance model, named iCAM, has been introduced. Via image appearance modeling new techniques for predicting overall image quality, without the need for intimate knowledge of the imaging system design, can be created.

1	INTRODUCTION.....	12
1.1	Device-Dependent Image Quality Modeling.....	13
1.2	Device-Independent Image Quality Modeling	15
1.3	Research Goals	17
1.3.1	Modular Color Image Difference Framework.....	19
1.3.2	Image Appearance and Quality Metrics	20
1.4	Document Structure	20
2	DEVICE-DEPENDENT IMAGE QUALITY MODELING.....	21
2.1	System Modeling	21
2.2	Subjective Quality Factor (SQF)	23
2.3	Square-root Integral (SQRI)	26
2.4	Device-Dependent Image Quality: Summary	28
3	DEVICE-INDEPENDENT IMAGE QUALITY MODELS: THRESHOLD MODELS .	29
3.1	Visible Differences Predictor (VDP).....	30
3.2	Lubin’s Sarnoff Model	37
3.3	Threshold Model Summary.....	40
4	DEVICE-INDEPENDENT IMAGE QUALITY: MAGNITUDE MODELS	42
4.1	S-CIELAB.....	42
4.2	Color Visual Difference Model (CVDM).....	47
4.3	Magnitude Model Summary.....	48

5	DEVICE-INDEPENDENT IMAGE QUALITY MODELING: COMPLEX VISION MODELS	49
5.1	Multiscale Observer Model (MOM).....	49
5.2	Spatial ATD	52
5.3	Summary of Complex Visual Models.....	53
6	GENERAL FRAMEWORK FOR A COLOR IMAGE DIFFERENCE METRIC	54
6.1	Framework Concept: Model Simplicity.....	54
6.2	Framework Concept: Use of Existing Color Difference Research.....	55
6.3	Framework Concept: Modularity	56
6.4	Framework Evaluation: Psychophysical Verification	59
6.5	General Framework: Conclusion	59
7	MODULES FOR IMAGE DIFFERENCE FRAMEWORK	60
7.1	Spatial Filtering Module	60
7.1.1	Barten CSF from Square Root Integral Model (SQRI).....	65
7.1.2	Daly CSF from the Visual Differences Predictor (VDP).....	66
7.1.3	Modified Movshon	67
7.1.4	Spatial Filtering Summary.....	68
7.2	Spatial Frequency Adaptation	69
7.2.1	Natural Scene Assumption	70
7.2.2	Image Dependent Spatial Frequency Adaptation	72
7.2.3	Spatial Frequency Adaptation Summary	73
7.3	Spatial Localization Filtering	74
7.3.1	Spatial Localization: Simple Image Processing Approach	74
7.3.2	Spatial Localization: Difference of Gaussian	74
7.3.3	Spatial Localization: Frequency filtering.....	75
7.3.4	Spatial Localization: Summary	77

7.4	Local and Global Contrast	77
7.4.1	Local and Global Contrast Summary	79
7.5	Error Reduction	79
7.5.1	Structured Data Reduction	81
7.5.2	Data Reduction Summary	83
7.6	Color Space Selection	83
7.6.1	IPT	84
7.6.2	Color Space Summary	85
7.7	Color Image Difference Module Summary	85
8	PSYCHOPHYSICAL EVALUATION.....	87
8.1	Sharpness Experiment	87
8.1.1	Spatial Resolution	87
8.1.2	Noise.....	87
8.1.3	Contrast Enhancement.....	88
8.1.4	Sharpening	88
8.1.5	Experimental Design	88
8.1.6	Sharpness Results	91
8.2	Contrast Experiment.....	100
8.2.1	Lightness Manipulations	100
8.2.2	Chroma Manipulation.....	100
8.2.3	Sharpness Manipulation	100
8.2.4	Experimental Conditions	100
8.3	Print Experiment	104
8.3.1	Print Experimental Setup.....	106
8.4	Psychophysical Experiment Summary.....	114
9	IMAGE DIFFERENCE FRAMEWORK PREDICTIONS	115
9.1	Sharpness Experiment	115
9.1.1	Baseline	116

9.1.2	Spatial Filtering.....	118
9.1.3	Spatial Frequency Adaptation	122
9.1.4	Spatial Localization	125
9.1.5	Local and Global Contrast Module	128
9.1.6	Cascaded Model Predictions	129
9.1.7	Color Difference Equations.....	130
9.1.8	Error Image Reduction	132
9.1.9	Metrics for Model Prediction	133
9.1.10	Further Model Analysis: Replacing Psychophysics?	141
9.1.11	Sharpness Experiment Conclusions	144
9.2	Contrast Experiment.....	145
9.2.1	Lightness Experiment.....	145
9.2.2	Chroma Experiment.....	146
9.2.3	Sharpness Experiment	147
9.2.4	Psychophysical Replacement?	148
9.2.5	Contrast Experiment Conclusions.....	149
9.3	Print Experiment Predictions.....	149
9.3.1	Sharpness Experiment	150
9.3.2	Graininess Prediction.....	152
9.3.3	Image Quality Experiment	155
9.3.4	Print Experiment Summary	156
9.4	Psychophysical Experimentation Summary.....	157
10	IMAGE APPEARANCE ATTRIBUTES	159
10.1	Resolution Detection	161
10.2	Spatial Filtering	162
10.3	Contrast Changes	163
10.4	Putting it Together: Multivariate Image Quality.....	163
10.5	Image Attribute Summary	164

11	ICAM: AN IMAGE APPEARANCE MODEL.....	166
11.1	ICAM Image Difference Calculations	173
11.2	ICAM Summary	175
12	CONCLUSIONS AND FUTURE DIRECTIONS	176
12.1	Summary of Dissertation.....	176
12.2	Scientific Contributions	177
12.3	Conclusions and Future Directions.....	178
12.3.1	Future Directions	180
13	REFERENCES	181
A.	PSYCHOPHYSICAL RESULTS.....	187
	Sharpness Experiment: Combined Results.....	187
	Sharpness Experiment: Cow Images	188
	Sharpness Experiment: Bear Images.....	189
	Sharpness Experiment: Cypress Images	190
	Sharpness Experiment: Cypress Images	191
	Contrast Experiment: Lightness Manipulation Z-Scores.....	192
	Contrast Experiment: Lightness Manipulation Z-Scores.....	193
	Contrast Experiment: Chroma Manipulation Z-Scores.....	193
	Print Experiment: Image QUALITY, Portrait, RIT Data.....	194
	Print Experiment: Image SHARPNESS, Portrait, RIT Data	194
	Print Experiment: Image GRAININESS, Portrait, RIT Data	195

Print Experiment: Image QUALITY, Portrait, Fuji Data.....	195
Print Experiment: Image SHARPNESS, Portrait, Fuji Data	196
Print Experiment: Image GRAININESS, Portrait, Fuji Data	196
Print Experiment: Image QUALITY, Ship, RIT Data	197
Print Experiment: Image SHARPNESS, Ship, RIT Data	197
Print Experiment: Image GRAININESS, Ship, RIT Data	198
Print Experiment: Image QUALITY, Ship, Fuji Data	198
Print Experiment: Image SHARPNESS, Ship, Fuji Data	199
Print Experiment: Image GRAININESS, Ship, Fuji Data	199
 B. PSEUDOCODE ALGORITHM IMPLEMENTATION	200

1 Introduction

The fundamental nature of image quality can be simultaneously considered obvious and obscure. When shown two images it is very easy for most people to choose the image they consider to be of higher quality. Often this is synonymous with choosing the image they prefer. Yet when asked to qualify why they made the choice, these same people often become silent. The choice was obvious, but why the choice was made often evades them.

The inability to even qualify our own preferences yields an interesting scientific challenge. How can we be expected to create a computational model capable of predicting the perception of quality when we can barely explain our own actions? Luckily, it is generally not necessary to have a complete understanding of our nervous system in order to make ourselves sit up in the morning. If that were the case, many people would never get out of bed. Likewise, if we make a decision regarding image quality often enough, and this decision is consistent, we can learn enough about our own actions to “get out of bed.”

Image quality modeling has been the focus of research over the course of many years. Engeldrum¹ offers an excellent review of many of the different techniques used in the design and evaluation of various modeling techniques. The general definition of image quality modeling is the creation of a mathematical formula that is capable of predicting human perceptions of quality.¹ Engeldrum describes two distinct approaches to this mathematical formulation, which he defines as the *impairment* and *quality* approaches. The impairment approach can be thought of as the measurement of the decrease in quality of an image from some reference or ideal image. This can be extended slightly by including a measurement of the increase in quality from a reference image if an ideal image does not exist. The quality approach, as defined by Engeldrum, attempts to model mathematically the quality of an image directly, without the need for a reference image. This can be thought of as comparing an image directly against some fundamental ideal mental representation.

These two fundamentally different approaches can be used in a similar context. The context put forward by Engeldrum, and shown in Figure 1, is the *Image Quality Circle*.²

Figure 1. Image Quality Circle, from <http://www.imcotek.com>

The Image Quality Circle illustrates the fundamental approaches that are generally used to tackle the problem of image quality modeling. The ultimate goal of any image quality modeling has been defined in this research as the ability to predict human perceptions of quality. This is illustrated in the top block of the quality circle, labeled “Customer Image Quality Rating.” To achieve this goal one can travel around the circle starting at any of the other blocks. In the context of the Image Quality Circle, there are two distinct approaches (or directions) to arrive at the destination, or goal. These two approaches have been described as *vision-based* or *systems-based*. To clarify the distinction between the two approaches to image quality modeling, Fairchild³ introduced terminology similar to traditional color imaging. He described the vision-based approach as *device-independent image quality* modeling, while the systems-based approach is described as *device-dependent image quality*.

1.1 Device-Dependent Image Quality Modeling

Device-dependent image quality modeling can be thought of as traveling throughout the right-hand side of the Image Quality Circle, as illustrated by the right-hand side of Figure 2.

Figure 2. Device-Dependent Paths for Image Quality Modeling

Essentially this approach attempts to relate systems variables (or technology variables as shown in the IQ circle), such as resolution, gamut volume, noise, MTF, and system contrast with overall image quality. This path would be equivalent to traveling the dark solid path shown in Figure 2. This path is often taken directly to describe the “quality” of an imaging system. This approach can be both valid and useful, when careful experimentation has been undertaken to link the system variables to the perception of quality. Often this approach can be misleading, as the system variable might have little effect on the overall human perception of quality. For example, saying “Printer X is twice as good as Printer Y, because it has 1200 dpi instead of 600 dpi” might not be valid if the system variable of *dpi* does not have a direct link to quality.

A more fundamental approach to systems-based modeling is to relate the system variables directly to perceptions using psychophysical techniques. This approach can be very successful when there is complete control of the imaging system. A very structured systems-based approach using these techniques is described in detail by Keelan and Wheeler^{5,6,7} This approach is described in further detail in Section 2.1.

The psychophysical experimentation necessary to relate the systems variables with human perception is often difficult to obtain. Several researchers have worked on creating models of the human visual system to replace the psychophysics. This is the approach taken by Granger’s Subjective Quality Metric (SQF).⁸ The SQF model uses properties of the imaging system as well as the human visual system to relate directly to the perception of quality. This approach is described in further detail in Section 2.2. A similar approach, though mathematically more complicated, was taken by Barten⁹ in the Square Root Integral (SQRI) model, as described below in Section 2.3.

One important consideration of these systems-based approaches to image quality modeling is the need to have intimate knowledge of the imaging system itself, either through the independent variables or through the system-wide MTF (modulation transfer function). This knowledge is most often available when designing imaging systems, which is where the systems-approach has traditionally been used with success. The device-dependent approach is not designed for predicting the quality, or quality difference, of any given image or image pair. To predict image quality regardless of the imaging device used to obtain the image requires the use of a device-independent image quality model.

1.2 Device-Independent Image Quality Modeling

Device-independent image quality models attempt to predict the human perception of an image without knowledge of the image origins. The most successful of these models are often referred to as perceptual models.^{1,3} The general idea is to model various aspects of the human visual system, and then to use these models to predict perceptual quality responses. This approach is slightly different from the vision modeling used in the device-dependent techniques described above. Whereas those models typically are used to create the link between system variables and quality, they are not generally concerned with directly modeling appearances. The vision modeling performed in device-dependent image quality is most often used to generate a single number that correlates system variables with overall quality. The vision models used for device-independent quality do not necessarily attempt to generate a single “unit of quality.” Often they are used to formulate individual image perceptions such as sharpness and colorfulness. These percepts are referred to as the “nesses” in the context of the Image Quality Circle. Device-independent image quality modeling falls into the left half of the Image Quality Circle, as shown in Figure 3.

Figure 3. Device-Independent Paths For Image Quality Modeling

The input for device-independent image quality modeling can be the image itself, or physical image parameters. These parameters can be measured aspects of the image, such as graininess and dynamic range for a hardcopy image. Likewise they can come from careful characterization of the viewing conditions for softcopy display. The relationship between the percepts such as contrast and sharpness with the image attributes can then be found using psychophysical experimentation. Ideally, models of the human visual system can be used to replace or supplement the psychophysics and predict the perceptual response to the measured image attributes. The individual percepts can then be combined into a general model of overall quality.

The techniques for device-independent image quality modeling can also be split into the fundamentally distinct approaches as described by Engeldrum, impairment and quality.¹ The *quality* approach attempts to model the judgment of image quality directly from the image itself. This might result from perceptually modeling the “nesses” of an image, perhaps compared to a mental interpretation of an ideal image. This approach might be considered the ultimate goal of image quality modeling, as the results might truly model the idealized perception of quality. Such an approach is difficult, as many aspects of quality can be considered very scene dependent. Many researchers have instead undertaken the *impairment* approach, in which image quality is modeled as a function of a reference image. If the reference image is the mental ideal image, then the impairment approach models a decrease in perceived quality and should be identical to the quality approach. If the reference is not the ideal image, then the impairment approach models the difference in quality, whether that change is an increase or a decrease.

One of the overall goals of image quality modeling is to eliminate the need for extensive human experimentation. It is doubtful that any model can ever replace psychophysical experimentation, so perhaps a more accurate goal would be to have the image quality models supplement and guide experimental design. In order to achieve this goal, it is necessary to model various aspects of the human visual system. As such, often these device-independent models are described as vision-based perceptual image quality models. These perceptual models can be further broken down into *threshold* and *magnitude* models.

Threshold models are typically generated using the impairment approach to image quality modeling. These models use the properties of the human visual system to determine whether or not there is a perceptible difference between two images. This “threshold” difference between two images is often called a Just Noticeable Difference (JND). It is important to note that a threshold model is incapable of determining the magnitude of the perceived difference, only whether there is a difference at all. Two such threshold models are Daly’s Visible Differences Predictor (VDP)¹⁹ and Lubin’s Sarnoff Model.²⁰ These models are discussed in detail in Sections 3.1 and 3.2 respectively.

While it is very beneficial to determine threshold differences between images, this is not sufficient for building an image quality model. In order to do this, one must also be able to determine the magnitude of the perceived differences (also called supra-threshold differences). Several vision-based models are capable of predicting magnitudes of differences. Perhaps the simplest of such models are the CIE color difference equations, such as CIE ΔE^*_{ab} . These equations are based on CIE colorimetry, which is designed to predict matches in simple color stimuli when viewed in a common condition. While the CIE system was designed specifically for simple color patches on a uniform background, many researchers have used the color difference equations as a type of image quality model. This approach is very limited, as it does not take into account many of the properties of the human visual system, such as spatial vision. Several researchers have extended the basic CIE system to include parameters for spatially complex viewing conditions. These extensions have been used to create more complete image quality models. Two such examples are the S-CIELAB³³ spatial extension to CIELAB, and the Color Visual Difference Model (CVDM).³⁴ These models build upon the CIE color difference equations to predict magnitude differences for complex image stimuli. S-CIELAB and the CVDM are discussed in detail in Sections 4.1 and 4.2 respectively.

The vision-based models introduced above take into account many properties of the human visual system, but generally do not try to model the exact physiology of the visual system. Rather, they can be considered empirical models of the visual system. Several more complicated vision-based models attempt to follow human physiology to a greater extent. These models do not rely on the CIE color difference equations for predictions. Generally, these more complex models are capable of predicting a wider range of spatial and color phenomena. Two such models are the Multiscale Observer Model (MOM)³⁹ and the Spatial ATD (Achromatic, Tritanopic, Deuteranopic) Model.⁴¹ These models are discussed in detail in Sections 5.1 and 5.2 respectively.

1.3 Research Goals

Of the two paths around the Image Quality Circle described above, the device-independent approach to image quality modeling appears to be a more generalized approach. The goal of an image quality model is to mathematically predict human perceptions of quality, so it seems necessary to incorporate a model of the human visual system into that model. As such, this research has focused on designing and evaluating a perceptual model capable of measuring image differences, quality, and appearance.

This research has generally focused on the *impairment* approach to image quality, as described by Engeldrum.¹ A description of this approach can be summarized as follows:

An image quality metric can be derived as a measure of the perceived quality difference from an ideal image.

Given an ideal image or an image of perfect “quality” then the impairment approach to image quality is quite plausible, as one simply needs to figure out how a given image varies from the ideal. This generalized concept is illustrated in Figure 4 adapted from Fairchild.⁴

Figure 4. Iconic Representation of Impairment Image Quality Modeling

The image on the left side of Figure 4 represents the “ideal” image. The ideal image can be thought to be some mental representation of a high-quality image. The other images can then be placed along the “scale” of image quality, with decreasing quality going to the right. The differences in each of the images can vary in dimension and type, as those shown represent many changes that might occur in an image reproduction system. The magnitude of the differences relates to the placement on the quality scale. (It should be noted that all the images represented are just iconic versions and are not meant to accurately reflect an actual quality scale.)

Using this impairment technique as a guideline, a metric capable of predicting perceived magnitude of differences between images goes a long way towards the ultimate goal of an image quality model. The first goal of this research is the formulation and evaluation of a color image difference metric. This metric is based on both spatial and color properties of the human visual system. As it is yet impossible to create a representation of the mental ideal image, calculating difference from a reference image is relied upon.

The CIE system of colorimetry and more specifically the CIELAB color space and associated color difference equations have proven successful in predicting perceived differences in simple color patches. Likewise, extensions to the CIELAB space, such as the S-CIELAB model have shown that the

color difference concept can be extended for use with spatially complex digital images in a relatively straightforward manner.³³ Therefore, it was hypothesized that an image difference metric capable of predicting both spatial and color differences can be built upon the CIELAB color difference equations. The resulting calculations create a spatially localized color difference map that should relate to overall perceived visual differences. This is, in effect, equivalent to sampling continuous scenes with a digital imaging device and determining perceived color differences at each sample point after removing those spatial aspects of the scenes that are below the limits of detection..

While a metric capable of predicting perceived color differences in images is valuable in its own right, it is not necessarily adequate for predicting overall image quality differences. Another important step towards an image quality model is to predict where the differences came from. Determining the magnitude and direction (or the cause and the amount) of the differences results in a measurement of overall perceived image appearance. This can be thought of as predicting the perceptual “nesses” such as sharpness and colorfulness, along with traditional color appearance correlates such as chroma, lightness, and hue. Thus, another goal of this research was the creation of a general form for an image appearance model using the image difference metric as a guideline.

1.3.1 Modular Color Image Difference Framework

The first stage of this research focused on the formulation of a general modeling framework for the creation of a color image difference metric. This framework can be thought of as a series of guiding principles and modeling techniques. The general framework allows for aspects of both spatial and color vision to be utilized in a single unified metric. The color image difference metric was designed with two important properties: simplicity and extendibility. If there is any hope of a color image difference metric gaining wide acceptance, the metric must (at the core at least) be relatively simple to understand, calculate and extend.

A simple model might be unable to predict the complex spatial properties of the human visual system. By creating a modular framework it should be possible to create a metric that begins as a simple core, such as the CIE color difference equations, and has more complicated building blocks that can be added when the complexity of the situation warrants. Modularity allows for any block in the framework to be removed, replaced, or enhanced without affecting any of the other blocks.

This concept might be better explained by thinking of the automobile industry. There is a general framework of an automobile: four wheels, roof, doors, engine, steering wheel, etc. Each of these objects represents an individual component, or module, of the automobile. When designing a new car, designers tend to follow the general “car” framework, but are allowed great flexibility in picking the individual

components that make up the car. Often the modular nature of the components allow for great flexibility as the choices are mostly independent of each other. For instance the choice of tires does not necessarily influence the choice of engine or body style. Sometimes the individual components do influence each other, as the choice of tires must be influenced by the size of the wheels. It is obvious by examining any busy parking lot that there are many different styles of cars to suit many individual necessities and tastes. Most of the automobiles are built with the same general “car” framework. The idea behind the modular color image difference framework is to allow for a similar freedom of choice and design.

1.3.2 Image Appearance and Quality Metrics

A color image difference metric can serve as a basis for the formulation of an image quality model. Another goal of this research has been the creation of a foundation upon which future computational image quality models can be built. This foundation combines aspects of the image difference metric with aspects of color appearance models to create an image appearance model. An image appearance model is capable of not only predicting image differences but also in predicting the general appearance of images (what causes the differences). An image appearance model should not be limited to the traditional color appearance correlates such as lightness, chroma and hue. Rather it should supplement those with image correlates such as sharpness, and contrast. These are often referred to as the “nesses” with regard to Engeldrum’s Image Quality Circle.²

1.4 Document Structure

This document details the steps followed to achieve the research goals described in the above sections. Further detail into existing device-dependent and device-independent image quality models will first be described In Sections 2-5. The lessons learned from these models were used to create the foundation for the color image difference metric. This modular foundation is described in detail in Section 7. Individual aspects and features of the color difference model are described in Section 8. As the goal of any image quality or image difference model is to mathematically predict human perceptions, the models must be evaluated against experimental data. Several psychophysical experiments used to design and evaluate the model are described in Section 9. The model predictions of these experiments are described in Section 10 and 11. The concept of an image appearance model, as well as an introduction of such a model is presented in Section 12.

2 Device-Dependent Image Quality Modeling

Device-dependent modeling concerns itself with the effect an imaging system has on the perception of overall image quality. There are two measurements needed for this type of quality modeling: measurement of system variables such as MTF, grain, addressability, etc., and measurement of perceived image quality obtained through the use of experimental psychophysics. Statistical methods can then be used to link the system variables to the psychophysically derived quality scales. With this type of modeling approach, it is often not necessary to understand the system variable being tested. This is equivalent to thinking “if I turn this knob on this machine, the image quality will change.” In the above example the quality scale is directly linked to the amount the “knob” is turned, without having to understand what turning the knob actually does to the output images.

Device-dependent modeling relies heavily on physical measurement of system variables as well as psychophysical experiments. This type of model has proven to be very successful for design and evaluation of many imaging systems. The following sections review one such unified approach to device-dependent modeling, described by Keelan^{5,6} and Wheeler.⁷ The need to perform extensive psychophysical experimentation to create the image quality scales used in this type of modeling might be seen as a drawback to these techniques. To reduce the need for exhaustive psychophysics several researchers have attempted to create a front-end model of the human visual system that combines with the modeling techniques of the imaging systems themselves. Two such models are Granger’s SQF metric⁸, and Barten’s SQRI metric.^{9,10}

2.1 System Modeling

Researchers in industry often have very precise techniques for device-dependent image quality research. This approach attempts to directly link perceptual quality to various imaging system parameters, such as MTF and noise. At the heart of this approach are the experimental techniques used to link system parameters to perceptions. One such approach involves the creation of an actual physical scale of image quality, referred to as the primary standard.⁶ This quality scale is created through extensive psychophysical evaluation, and is designed to have meaningful units of measurement. Keelan describes several techniques that can be used to create (or recreate) the standard scale.⁵ The primary standard can then be used to link different imaging system variables back to the perceptual scale of image quality. This linking is done by first creating a series of images that vary across a system parameter. These images are then compared against the primary standard, to create a relative scale of quality for that particular system variable. This can be repeated for many different parameters, creating a series of quality scales for each

parameter. These individual scales can then be combined to create an overall metric of perceived image quality.

An example of this type of systems modeling is as follows. Suppose a researcher is interested in the effect of additive noise in an imaging system on perceived image quality. The researcher must first perform psychophysics on many images that contain additive noise, and link the results of the psychophysical test with the primary standard scale. One experimental method that is often used is a hardcopy quality ruler.^{5,8} A quality ruler is physical representation of the primary scale, generally created with series of images of “known” quality. These images are systematically varied such that they span a wide range of quality, in uniform steps of known value. In order to create the ruler and link it to a known scale, exhaustive psychophysics are necessary. Several psychophysical techniques for creating a precise ruler are well described by Keelan.⁵ It should be noted that the systematic variable scaled and represented on the ruler does not necessarily have to be the same as the variable being tested. One type of systematic variation that is well controlled and quantified is that of changing the Modulation Transfer Function (MTF) of the imaging system. This has the effect of altering the spatial frequency of the reproduced images. A quality ruler with simulated variations in MTF is shown below.

Figure 5. Image Quality Ruler with Simulated Variations in MTF

The quality ruler is then used to as a reference to scale the system variable that is of concern, such as additive noise in this example. An observer is asked to judge the quality of a given image by matching it with the quality of one of the ruler images. Since the quality of the ruler images is well quantified, the quality of the image being judged is now equally quantified, presuming all the observers map the new image in a similar way.

To produce an objective metric it is necessary to be able to measure the variations in the images themselves. For the additive noise example one possible measurement could be to measure the Wiener Noise Spectrum.¹¹ This measurement can then be directly related to the quality scale. For all future versions of this imaging system, it is not necessary to repeat the above experiment to produce scales of quality. Instead, it is only necessary to measure the objective metrics and use the relationship between that metric and the quality scale, assuming all other viewing conditions are constant. This technique has been proven successful for many different types of system variables.⁵ It is important to note the limitations of

this technique in regards to scene dependency. Different scene content might result in different quality scales, especially if the images used in the quality ruler differ from those used for the scaling experiment.

The goal of this type of systems modeling is to create objective metrics that correlate strongly with a known scale of image quality. If the known scale is in turn linked to a single “primary standard” it is also possible to link different objective scales together into a single multivariate metric of image quality. One method that has been used to link uni-variate metrics together into a single model is using the Minkowski metric, first used for image quality research by Bartleson¹² and shown below:

$$\Delta Q = \left(\sum_i (\Delta Q_i)^r \right)^{1/r} \quad (1)$$

where ΔQ is overall change in quality, ΔQ_i is the change in quality resulting from any given objective or attribute, and r is the power of the Minkowski metric. The value of r can be optimized to fit an existing set of objective attributes. It should be noted that when $r=2$ the Minkowski metric reduces to a standard Root Mean Square (RMS) equation which can be thought of as a Euclidian distance metric where the objective attributes are independent or orthogonal.

It should also be noted that this type of device-dependent modeling could be used to link a system variable directly with human perceptual attributes, or the “nesses.” For example, MTF can be linked to the perception of sharpness, or gamut-volume with colorfulness. In this situation, the objective metrics become the appropriate “ness.” A total image quality model can then be created out of the individual perceptions, perhaps using the Minkowski summation technique to rank the importance of each ness to overall quality.

This type of device-dependent modeling has proven to be a successful method for quantifying quality for many different imaging systems.^{1,5,6,7,12} The weaknesses in this type of model are the need for extensive psychophysical experimentation, both for the creation of the primary standard, and then for each attribute (such as noise, resolution, etc.) being scaled. Replacing the psychophysics and objective modeling with a single model combining elements of the human visual system with the imaging system variable is the goal of other device-dependent models, such as the SQF and the SQRI.

2.2 Subjective Quality Factor (SQF)

Granger and Cupery introduced a Subjective Quality Factor (SQF) in an effort to combine the properties of the human visual system (HVS) with the optical properties of any given imaging system to get a combined metric of image quality.⁸ This metric begins with the Optical Transfer Function (OTF) of an imaging system. An imaging system OTF can be divided into magnitude (Modulation Transfer Function, MTF) and phase transfer functions (PTF) and represents how well a system reproduces information at any

given spatial frequency. As such these functions are often used to represent system performance. Granger and Cupery first hypothesized that the area under an OTF curve would work well as an objective metric for perceived image quality of a system, as that area determines the overall amount of information an imaging system is capable of reproducing. An example of this type of metric is shown below:

$$Q = \int OTF(f) df \quad (2)$$

where Q relates to quality, $OTF(f)$ is the optical transfer function of the imaging system, and f is spatial frequency. It was recognized that this metric must take into account the properties of the human visual system, specifically the Contrast Sensitivity Function (CSF) of the human eye. The CSF was referred to as the MTF of the human eye, though this can be misleading as the CSF takes into account more of the human visual system than just the optics of the eye. An example of this is shown in Figure 6.

Figure 6. Example of MTF of the human visual system.

The quality factor could then be rewritten by cascading the MTF of the human visual system as shown below.

$$Q = \int OTF(f) \cdot MTF_{eye}(f) d(\log f) \quad (3)$$

It is important to note that this assumes linearity of both the imaging system and the human visual system. While it is known that the visual system often behaves in a nonlinear manner, this assumption allows for simplicity in the modeling. To accommodate the nonlinear nature of the human visual system the SQF

used the logarithm of the integrated area, as shown in Equation 3. In Equation 3 the limits of integration are set by the overall frequency response of the human visual system, the MTF. Taking this one step further, Granger and Cupery recognized the band-pass nature of the human visual system, and sought to simplify the quality factor equation by limiting the MTF to a rectangular band-pass function, where the MTF is zero below and above certain frequencies and one between them as shown in the shaded area of Figure 6. The quality factor equation now reduces to:

$$Q = \int_{f_1}^{f_2} OTF(f) d(\log f) \quad (4)$$

where f_1 and f_2 represent the limits of integration as defined by the frequency band-pass of the visual system, originally defined as 10 and 40 cycles/mm. This band-pass region was originally scaled to the retina of the observer, and included the magnification of the imaging system.⁸ This region corresponds to 3 and 12 cycles per degree of visual angle for a typical viewing distance. At the time, this simplification greatly eased the calculation cost of the SQF.

The quality factor expressed in Equation 4 utilizes a one-dimensional OTF for the imaging system, and band-pass for the visual system. The complete version of the SQF extends this assumption to two-dimensions, for use with actual images. This final equation is shown below:

$$SQF = \int_{f_1}^{f_2} \int_0^{2\pi} OTF(f, \theta) d(\log f) d\theta \quad (5)$$

where f is the spatial frequency for a given line structure along an azimuth angle θ . An important consideration is the logarithm factor in the SQF, represented by $d(\log f)$ in Equations (2-5). This logarithm factor can be thought of as a multiplication of the MTF by a $1/f$ factor in the frequency domain, essentially performing an integration over the visual field in the spatial domain.¹³ This can be explained by noting that $d(\log f) = df/f$. This $1/f$ weighting will be revisited in the spatial frequency adaptation discussed in Section 7.2.1.

The SQF has been shown to predict image quality for several experimental conditions.^{8,14, 15} Essentially the SQF was capable of replacing both the psychophysical experimentation and objective function definition described in the above systems modeling section.

The SQF has proven to be successful in the prediction of several different experimental datasets. This is especially impressive considering the simplicity of the formula itself. This type of device-dependent model has several weaknesses, though. The model works on the assumption of predicting the image quality capabilities of an imaging system based entirely on the OTF of the imaging system itself. This means that the model is only capable of predicting quality loss caused by the OTF, and not by other factors such as dynamic range and gamut volume. This type of modeling is also limited to the prediction

of an entire end-to-end imaging system, and must be recalculated if any of the individual components are changed. The SQF also ignores any color information, essentially assuming that only the luminance channel is a factor in image quality. It has also been shown that the simple rectangle shaped band-pass assumption of the MTF of the visual system can be an over-simplification, especially over a wide range in viewing conditions.¹⁶ Barten has specifically addressed this last weakness in the SQRI model.⁹

2.3 Square-root Integral (SQRI)

The Square-root Integral (SQRI) equation builds upon the SQF metric by taking into account some of the nonlinear behavior of the human visual system, as well as adding a more complex model of the CSF of the human visual system.⁹ The complex CSF takes into account many of the viewing conditions, where the SQF did not. In addition the square-root compression models more of the nonlinear behavior of the human visual system than the SQF model.⁹ The contrast sensitivity function is used as a threshold modulation function, so that a just-noticeable difference (JND) at any given spatial frequency is equalized. The final form of the SQRI model is shown below:

$$SQRI = \frac{1}{\ln 2} \int_{f_{\min}}^{f_{\max}} \sqrt{\frac{MTF(f)}{MTF_{eye}(f)}} \frac{df}{f} \quad (6)$$

where f is the spatial frequency in cycles per degree of visual angle, $f_{\min}$ and $f_{\max}$ are the minimum and maximum spatial frequencies resolvable by the visual system, MTF is the modulation transfer function of the imaging system, MTF_{eye} is the threshold modulation of the visual system, and df/f is the logarithmic integration over spatial frequency, recalling that df/f is equivalent to $d(\log f)$. The MTF_{eye} is used as a model of the contrast sensitivity function of the human eye. The function used by Barten is a complex model that takes into account many viewing condition factors. The full equation of this is shown below.

$$CSF(f) = \frac{1}{MTF_{eye}(f)} = \frac{e^{-2\pi^2 \sqrt{\sigma_o^2 - (C_{ab}d)^2} f^2}}{k \sqrt{\frac{2}{T} \left(\frac{1}{X_o^2} + \frac{1}{X_{\max}^2} + \frac{f^2}{N_{\max}^2} \right) \left(\frac{1}{\eta p E} + \frac{\Phi_o}{1 - e^{-(f/f_o)^2}} \right)}} \quad (7)$$

The various factors can be altered depending on many aspects of the viewing conditions, including adapting luminance (E), viewing distance (d), image size (X_o), and viewing time. (T). For typical values, consult Barten.¹⁰ An example of the shape of this function for a typical viewing condition of 100 cd/m² is shown in Figure 7. The parameters used to create Figure 7 are the “default” values, and are defined in Table 1.

Example Barten CSF

Figure 7. Barten CSF for an average condition.

Table 1. Example Parameters for Barten CSF

CSF Parameter	Value
σ_o	0.50
C_{ab}	0.08
k	3.00
T	0.10
X_{max}	12.00
N_{max}	15.00
η	0.03
ρ	1.2×10^6
Φ_o	3×10^{-9}
E	1200
X_o	10.00
d	4.00
f_o	7.00

The SQRI can be extended to two-dimensions similarly to the SQF. It is important to note that the 2-D CSF function is assumed to be isotropic, *i.e.* not orientation specific.

The combination of the more precise modeling of the human visual system, in addition to the nonlinear square-root function has proven to be successful at predicting a variety of image quality experiments.^{10,17,18} As such, it can be used to replace the psychophysical experimentation as described in the above systems method. The similarity to the SQF metric leaves the SQRI model with the same inherent weaknesses. This model can only be used in conjunction with the MTF of an entire system. This must be carefully measured for any hope of a good prediction, and must be changed if any component of the imaging system is altered. Similarly, the SQRI model completely ignores any color information, thus assuming all image quality judgments rely only on luminance information.

2.4 Device-Dependent Image Quality: Summary

This section outlined several experimental techniques and procedures that have been shown to be successful predictors of image quality. These techniques can be described as device-dependent predictors of image quality, as they attempt to directly link imaging system parameters with human perceptions. In order to fully utilize these techniques it is important to have intimate knowledge of the imaging system used to create the images. When that knowledge is available, these device-dependent techniques can be very powerful tools for measuring and predicting overall quality. Sometimes knowledge of the imaging system is difficult or impossible to obtain. In these cases, it is often desirable to predict image quality from the images themselves. The techniques used to do this make up device-independent image quality metrics. Several such techniques are described in subsequent sections.

3 Device-Independent Image Quality Models: Threshold Models

Another approach for image quality modeling replaces knowledge of the performance of the imaging system itself with actual images as input into the models. Since there is no need to have any knowledge of the image origin, this approach is referred to as device-independent image quality modeling. Several common approaches to this type of modeling use knowledge of the human visual system to in conjunction with the image data.¹ Because the input to these models is the images themselves, this can also be thought of as an image-processing approach to image quality. The benefits of this technique for quality modeling over a device-dependent vision model, as described above, can be quite significant. There is no need to fully characterize all elements of an imaging chain, as the effects of those elements are contained in the images themselves. Using the image as input into the model also preserves all phase information of the image, which is lost when describing an entire imaging chain with a single function such as an MTF. Phase information can be an important aspect in visual perception of complex stimuli, as demonstrated by the phenomenon of visual masking.¹⁹ In certain situations specific spatial frequencies might not be visible depending on their location in the image.

Image-based vision models typically work as relative image quality models. They do not give an absolute value of image quality, but rather can determine the difference in image quality between an image pair. This approach can be considered the impairment approach to image quality, when used with a standard reference image. The first stage in determining overall differences in image quality between two images is determining if there is any perceptible difference between the images. This can be thought of as a detectability metric, or a threshold model. Two such models are Daly's Visible Differences Predictor (VDP)¹⁹ and Lubin's Sarnoff Model.^{20,21} These models are both threshold vision models, in that they deal with the probability of subthreshold detection, rather than the magnitude of suprathreshold detection. This type of model can be very important for imaging systems design. For instance, when designing a new type of image compression algorithm it can be beneficial to have a model based on human perceptions to determine if the compressed image looks different than the original. The alternative would be to design a psychophysical experiment every time a change is made to the compression algorithm.

While not actually capable of providing an absolute metric for image quality, these threshold models do provide an important step in that direction. For that reason, they will be discussed in further detail. The VDP and the Sarnoff model are similar in form, with several distinct differences. These differences are described in more detail below.

3.1 Visible Differences Predictor (VDP)

The Visible Differences Predictor is an image-processing model based on properties of the human visual system. It is designed to predict the probability of detection of differences between two images.¹⁹ The general form of the model is illustrated below in Figure 8.

Figure 8. Flow chart of Visible Differences Predictor.

The VDP takes two images as input, called the original and reproduction for these purposes. Also input into the model are several factors that are related to the physical viewing conditions, including luminance of the images in cd/m^2 , image size, viewing distance, and viewing eccentricity. The input images are first transformed via a local amplitude nonlinearity. This is an attempt to reconcile the fact that the HVS

perception of lightness is nonlinearly related to luminance. For the purposes of the VDP, the luminance adaptation of any given pixel is determined by only the luminance of the pixel itself.¹⁹ The capability to perform a local contrast adaptation exists in the VDP, which is illustrated by the family of nonlinear curves in Figure 8. The choice of curves used can depend upon the average luminance of the neighboring pixels, if desired. The adjusted “lightness” images are then modulated by the Contrast Sensitivity Function (CSF) of the human visual system. This is very similar to the process of spatial filtering in the SQF and SQRI model.

The CSF function used in the Visible Differences Predictor is very complete, and capable of predicting the effects of many changes in the viewing conditions. These changes include the broadening and flattening of the CSF as a function of luminance level, as well as compensations for viewing distance and image orientation. The functional form of the CSF is show below:

$$CSF(f, l, i^2) = \left(\left(3.23 (f^2 i^2)^{-0.3} \right)^5 + 1 \right)^{-0.2} A_1 \epsilon f e^{-(B_1 \epsilon f)} \sqrt{1 + 0.06 e^{B_1 \epsilon f}} \quad (8)$$

$$A_1 = 0.801 \left(1 + \frac{0.7}{l} \right)^{-0.2} \quad B_1 = 0.3 \left(1 + \frac{100}{l} \right)^{0.15}$$

where f is spatial frequency in cycles per degree of visual angle (cpd), i^2 is image area (assuming square image), ϵ is a frequency scaling constant, and l is the light adaptation level in cd/m^2 .

This function can be further extended to add orientation selectivity by accounting for accommodation level, eccentricity, and orientation, as follows:

$$f(d, e, \theta) = \frac{f}{r_a r_e r_\theta} \quad r_a = 0.856 \cdot d^{0.14} \quad (9)$$

$$r_b = \frac{1}{1 + 0.24e} \quad r_\theta = \frac{1 - 0.78}{2} \cos(4\theta) + \frac{1 + 0.78}{2}$$

where d is viewing distance in meters, e is eccentricity (shift off foveal center) in degrees of visual angle, and θ is orientation in degrees. The 0.78 term in the r_θ term corresponds to the amount of modulation in the diagonal directions, essentially controlling the amount of change from an isotropic function.

The VDP contrast sensitivity function for typical viewing conditions of 100 cd/m^2 at 0.5 meters is shown in Figure 9. The viewing distance is important for the VDP because it changes the shape of the contrast sensitivity function due to changes in pupil-size and focal length, as described by the r_a term in Equation 9.

Figure 9. Typical Shape of VDP Contrast Sensitivity Function

The one-dimensional projection of the contrast sensitivity function as shown in Figure 9 looks very similar to that of the Barten SQRI model, as shown in Figure 7. While true on the horizontal and perpendicular axis, this is not the general case. The Barten CSF is completely isotropic, whereas the VDP uses an anisotropic orientation specific CSF. This can be seen in the two-dimensional CSF plot shown below in Figure 10. Essentially this accounts for the human visual system's decrease in sensitivity for off-axis (horizontal and vertical) stimuli. This is referred as the oblique effect, and has been shown to be an important feature of the visual system.^{22,23}

Figure 10. Two-dimensional VDP Contrast Sensitivity Function

The usage of the CSF in the VDP is similar in nature to the use in the SQRI model. The CSF function should not be considered the MTF of the human visual system, as that implies that the visual system is a linear system. Rather, the CSF is described as a threshold normalization function. The CSF is used as a linear filter to normalize all spatial frequencies and orientations such that the threshold for detection is identical. Essentially the CSF modulates high-frequency information more than low-frequency information. This means that in order to be detected, a difference at a high-frequency must be larger than a difference at a lower frequency. This is an important concept for the use of a CSF in an image-processing model. The Daly CSF represents a jump in complexity from the SQF and SQRI models. Section 9.2.1.2 discusses the relative performance of the different models in order to determine the importance of the added complexities.

The spatially filtered images are next fed into the detection mechanism of the VDP. The detection mechanism itself consists of four components: spatial frequency hierarchy, visual masking functions, psychometric functions, and probability summation. The result of these steps is a detection map, which can also be referred to as a threshold detection image.

The first stage of the detection mechanism is to decompose the input images into several distinct frequency bands. This corresponds to the frequency selectivity process thought to occur in the human visual system. The spatial-frequency decomposition is performed using a discrete Cortex transform.²⁴ The Cortex transform can be thought of as a discrete approximation and simulation of cortical receptive fields. The general goal of the transform is to decompose a given image into a set of images that vary in both spatial frequency content and orientation.²⁴ This is accomplished by first transforming the images into the frequency domain with a Fourier transform. The frequency images are then filtered with a series of frequency selective bands. The structure of these bands is illustrated in Figure 11, where the shaded area represents one set of filters (at a given frequency range and orientation). The filtered images are then transformed back into the spatial domain with the inverse Fourier transform.

Figure 11. Structure of Cortex Transform.

The frequency bands are created using a modified difference of Gaussians, called a Difference of Mesa (DOM) filters. The DOM filters are then separated into orientation bands called *fan* filters. The form of the Cortex Transform, along with the corresponding equations, is described in detail by Watson²⁴ and the modifications are described by Daly.¹⁹ At the center of the filters is the base filter, which is a modified single Gaussian. This serves as an isotropic low-pass function, and can be used for the local contrast calculations if desired. An example Cortex filter set corresponding to a base filter, four frequency channels (DOMS) and three fan orientations are shown in Figure 12.

Figure 12. Example of Cortex Transform Filters.

The top row of the Figure 12 represents the low-pass “base” filter. The following rows represent different ranges in spatial frequency, starting with high-frequencies and decreasing in range for each row. The columns represent different orientation selectivity. These filters are applied to an image by taking the Fourier transform of the image and multiplying by the series of filters and then taking the inverse Fourier transform back into the spatial domain, as shown below:

$$\text{Image}_k = \mathcal{F}^{-1}(\mathcal{F}(\text{image}) \cdot \text{Cortex}_k) \quad (10)$$

where $\mathcal{F}$ and $\mathcal{F}^{-1}$ represent the forward and inverse Fourier transform respectively, Image_k represents the individual sub-band image, and Cortex_k represents the individual Cortex Transform filter. The result of this transformation is a series of images, as shown in Figure 13. It should be noted that the Cortex transform, and Daly’s use of it is only described for the luminance channel. Thus the VDP, like the SQF and the SQRI is defined to ignore color information.

Figure 13. Sub-band Images Resulting From Cortex Transform.

Each of the sub-bands contains the information for a specific range of spatial frequencies and orientations. The individual sub-band images are then used to predict visual masking.

Masking, as a general term, refers to the effect on visibility one particular visual stimulus or pattern has on any other pattern or stimulus due to the general information surrounding the stimulus.^{25,26,27} Specifically in the VDP, masking refers to the decreased visibility of a stimulus due to the presence of a suprathreshold background.¹⁹ This is accomplished by examining the information contained in the individual bands relative the information contained either in the low-pass band or with the mean of all bands. This relationship is referred to as the contrast difference of the individual band, and is used as input to a visual masking function. The VDP is designed such that the contrast metric can be either a localized contrast that is based on a pixel-by-pixel location of each band with the corresponding pixel location of the base-band, or a global contrast that is relative only to the mean value of the base-band. The

result of the visual masking is the creation of threshold-elevation maps for each sub-band. Each elevation map is purely a function of the contrast in each sub-band.

The threshold-elevation map along with the global contrast difference for each of the input images are calculated. This information is then used to calculate a probability of detection for each sub-band. The general form of this probability is a psychometric function, and is as follows:

$$P_k[x, y] = 1 - e^{-\left(\frac{\Delta C_k[x, y]}{(T_{em}[x, y] - T)}\right)^\beta} \quad (11)$$

where x, y are the pixel locations in the image, P_k represents the probability of detection of the k^{th} sub-band, ΔC represents the contrast difference for a given pixel in a given sub-band, T_{em} is the threshold elevation mask, T is the pre-defined threshold, and β describes the slope of the psychometric function. The CSF filtering assures that the threshold, T , of each sub-band is the same for all frequencies. The probability of detection for all sub-bands is then combined to produce an image that contains the overall probability of detection at any given pixel. This combination is performed using a probability summation as follows.

$$P[x, y] = 1 - \prod_k (1 - P_k[x, y]) \quad (12)$$

The output of the VDP is thus an image that contains the probability of detection of difference at any given pixel. This can be thought of as either an error map, or error image. It is an important “feature” of the model that it does not reduce the error information into a single number, although this can easily be done using a simple statistical approach. Instead, it illustrates where errors occur, and allows the user the flexibility of determining what causes the errors, where they show up, and what to do to “fix” the system.

The VDP has proven to be capable of predicting the visibility of errors between any two given images.^{19,28,29} This can be very beneficial when designing and testing an imaging system. The VDP does have some weaknesses. The first weakness is complexity. It is a difficult model to implement since there are many free parameters involved that need to be chosen for a given situation. While this allows for greater flexibility, it also makes it difficult to choose the correct settings for a given situation. In addition, the VDP relies solely on luminance information and chooses to ignore color information. The color channels are capable of strongly influencing perceived differences in quality.³⁰

3.2 Lubin's Sarnoff Model

Another threshold difference model that is based on the human visual system is Lubin's JND Sarnoff model.²⁰ This model is similar in structure to the VDP, with several distinct variations. The general model is shown in Figure 14.

Figure 14. General Flow Chart of Sarnoff JNDmetrix from <http://www.jndmetrix.com>

The JND Sarnoff model begins with two RGB color images. Figure 14 shows four color images. These images are then subjected to “Front End Processing.” This is a combination of the visual system optics and sampling.²⁰ The details of this stage are proprietary in nature, but are described as a blurring of the images as a function of viewing distance using a 2D convolution kernel. As such, it can be considered similar in nature to the CSF filtering in the previously described vision models. The images are then resampled to model the sampling of the photoreceptors in the retina. This involves sampling with a square grid of approximately 120 pixels per degree. This grid creates a modeled retinal image of 512x512 pixels.²⁰

The retinal image is then converted into an opponent color space. The chosen space is luminance value Y , and CIELUV u^* v^* opponent coordinates. This transformation is accomplished through characterization of the display device. The luminance and chrominance channels are then converted to band-pass contrast responses. This is accomplished using an image processing technique called the Laplacian pyramid decomposition.³¹ The Laplacian pyramid is similar in nature to the Cortex transform, except it is based on efficient image processing computation rather than cortical simulation. The pyramid is a series of spatially low-passed images (Gaussian filtered). Each image in the series is limited to half the maximum spatial frequency of the previous image. A Laplacian pyramid of depth 4 is shown below in Figure 15. The Laplacian pyramid allows for efficient calculation of band-pass images, or “quasi-subbands,” using only simple convolution and image subtraction.³¹ It is similar to creating frequency band images, without the need to go into the frequency domain.

Figure 15. Example of Laplacian Pyramid of Four Levels.

Band-pass representations of the images can be found by subtracting the various low-pass images, as shown in the bottom row of Figure 15. This can be considered the same idea as the DOM band-pass filters in the Cortex transform, without the spatial orientation of the fan filters. The JND model typically uses a pyramid of depth seven, with the band-pass images being the difference of every other step rather than with each adjacent step. The band-pass images are centered in the range from 0.5 cpd to 32 cpd with a single octave between them.²⁹ The overall band-pass can be considered 2 octaves wide. The resulting Laplacian bandpassed images can be considered local difference images. When the bandpass images are normalized by the mean of the image the result in essence local-contrast measure, similar to the calculation of Weber contrast: $(L_{\max} - L_{\min}) / L_{\text{mean}}$.²⁰

To obtain the orientation-specific results the luminance pyramid is further filtered with the steerable filters of Freeman and Adelson.³² The specific filters used are the second derivative of a Gaussian and its Hilbert transform pair (a filter with the same frequency response, but with a phase shift of 90°). Four orientations are chosen resulting in 8 filtered images. These responses are used to calculate the orientation contrast by squaring and summing each Hilbert pair. The contrast values are then normalized by multiplying by the Barten CSF from the SQRI model (Equation 7) for each bandpass level in the pyramid. The visual masking of each orientation channel is performed using a nonlinear sigmoid transducer for each of the luminance and chrominance pyramid contrast channels. These transducers convert the contrast values into a contrast distance. This distance is then converted into a probability of detection using a psychometric function similar to that used in the VDP. The probability of detection can then be converted into a threshold JND.

It is important to note that the probability of detection is designed to be a single number, often the maximum of the spatial probability map. This single number is only designed to determine whether a human can see the difference between the two images. The model has sometimes been used to represent the “number of just noticeable differences” between the two images, although it was not designed as such. Care should be taken when using a threshold detection metric to produce magnitude scales of JNDs. It is also important to note that the Sarnoff JND model is proprietary in nature so many of the specific implementation details are not available, though it is possible to license the algorithm or purchase a hardware implementation from <http://www.jndmetrix.com>.

3.3 Threshold Model Summary

The Visible Differences Predictor and the Sarnoff JND metric are both comprehensive examples of device-independent image quality models. These models take into account many properties of the human visual system, so they can be considered perceptual models. They are designed to predict whether

there is a perceptible difference between a pair of images. When used with a standard reference image this is an example of the impairment approach to image quality modeling.

One possible benefit of using this type of perceptual model might be the replacement of exhaustive psychophysical experimentation. With a computational model that is capable of predicting the same results as a human observer, the need for extensive psychophysics is reduced. It is doubtful that such a computational model will ever totally remove the need for human observers, though it is hoped that such a model could help aid the design of experiments. For instance it might be possible to use such a model to determine the probability of detection of image differences when designing a threshold image experiment. If the probability of detection is too high or too low then it might not be necessary to have human observers view those specific image pairs. More experimental effort can be concentrated towards images close to the threshold.

The two models described in this section are similar in structure, though the VDP ignores color while the Sarnoff JND model is capable of utilizing color information. These models are designed to predict the “just noticeable differences,” often called threshold differences, between image pairs. These models were not designed to predict the magnitude or direction of these differences. Therefore, they are not capable in determining whether the difference is an error or an enhancement. Magnitude differences, often called supra-threshold differences, quantify the size of the difference between two images. In order to create a full image-quality scale it is necessary to be able to predict the magnitude of the error, as well as the direction. For models capable of predicting magnitude differences, a different class of device-independent image quality model is necessary.

4 Device-Independent Image Quality: Magnitude Models

We have already discussed how using the strengths and limitations of the human visual system can be beneficial in image quality modeling. The device-dependent approaches that utilized properties of the human visual system, as discussed in Sections 2.2 and 2.3 have shown to be capable of predicting many different types of psychophysical data.^{14,15,17,18,22,29} This device-dependent approach requires the understanding and characterization of the entire imaging system. Many times this information is unavailable. In these situations, it is desirable to have a model capable of predicting quality when given only images as input stimuli. Several perceptual threshold models that use images themselves as input were discussed in the previous section. These models are very capable of predicting whether there will be a perceptible difference between an image pair, but are generally not designed to predict the magnitude, or direction, of these differences. If there is to be any hope of predicting scales of image quality, then it is necessary to first be able to predict the magnitudes of image differences. In this section two such models that are capable of predicting magnitude differences, S-CIELAB³³ and the Color Visual Difference Model (CVDM)³⁴ are examined.

4.1 S-CIELAB

S-CIELAB was designed specifically as a spatial extension of the CIELAB color difference space.³³ The goal was to build upon the successful CIE color difference research, and produce a metric capable of predicting the magnitude of perceived differences between two images. The spatial extension is essentially a vision-based preprocessing step on top of traditional CIE colorimetry, and can be thought of as a spatial-vision enhancement to a color space and its corresponding color-difference equation. The source code to S-CIELAB is freely available on the internet.³⁶ The general model flowchart is shown in Figure 16.

Figure 16. General Flow-chart of S-CIELAB Model.

S-CIELAB takes as input an image pair, called the original and reproduction for this example. The images are then transformed into a device independent color representation, such as CIEXYZ or LMS cone responses. The primary advantage S-CIELAB offers over a standard color difference formula is the spatial filtering pre-processing step. This filtering is performed in an opponent color space, containing one luminance and two chrominance channels. These channels were determined through a series of psychophysical experiments testing for pattern color separability.³⁵ The opponent channels, AC_1C_2 , are a linear transform from CIE 1931 XYZ or LMS as shown below.

$$\begin{bmatrix} A \\ C_1 \\ C_2 \end{bmatrix} = \begin{bmatrix} 0.990 & -0.106 & -0.094 \\ -0.669 & 0.742 & -0.027 \\ -0.212 & -0.354 & 0.911 \end{bmatrix} \begin{bmatrix} L \\ M \\ S \end{bmatrix} \quad \begin{bmatrix} A \\ C_1 \\ C_2 \end{bmatrix} = \begin{bmatrix} 0.297 & 0.72 & -0.107 \\ -0.449 & 0.29 & -0.077 \\ 0.086 & -0.59 & 0.501 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} \quad (13)$$

Original Image

Achromatic (A)

Red-Green (C1)

Blue-Yellow (C2)

Figure 17. Opponent Color Representation of AC_1C_2

One important note about the AC_1C_2 opponent color space is that the three channels are not completely orthogonal. The chrominance channels do contain some luminance information, and vice-versa. This is illustrated in Figure 17, as the “white” lighthouse contains additional chroma information in both the red-green channel, and the blue-yellow channel. The lack of orthogonality has the potential to cause problems with the spatial filtering. For example, color fringing may occur when filtering an isoluminant image, since the achromatic channel contains some color information. It is important to stress that the opponent channels were based on a small number of actual observers, and were not designed specifically for image processing.

After both images are transformed into the opponent color space, the independent channels can be spatially filtered, using filters that approximate the contrast sensitivity functions of the human visual system. Three different filters are used, representing the different sensitivities of the three channels. The filtering is accomplished using either a series of convolutions in the spatial domain, or by using linear filtering in the frequency domain. The original S-CIELAB specification uses two-dimensional separable convolution kernels. These kernels are unit sum kernels, in the form of a series of Gaussian functions. The unit sum was designed such that for large uniform areas S-CIELAB predictions are identical to the corresponding CIELAB predictions. This is important, as S-CIELAB thus reduces to traditional color difference equations for large patches. The equations below illustrate the spatial form of the convolution kernels:

$$filter = k \sum_i w_i E_i \quad (14)$$

$$E_i = k_i e^{-\frac{x^2 + y^2}{\sigma_i^2}} \quad (15)$$

The parameters k and k_i normalize the filters such that they sum to one, thus preserving the mean color value for uniform areas. The parameters w_i and σ_i represent the weight and the spread (in degrees of visual angle) of the Gaussian functions, respectively. Table 2 shows these values for the kernels used in S-CIELAB. It is important to note that these values differ slightly from the published values, as they are already adjusted to sum to one.³⁶

Table 2. Weight and Spread of Gaussian Convolution Kernel

<i>Filter</i>	<i>Weight (w_i)</i>	<i>Spread (σ_i)</i>
Achromatic 1	1.00327	0.0500
Achromatic 2	0.11442	0.2250
Achromatic 3	-0.11769	7.0000
Red-Green 1	0.61673	0.0685
Red-Green 2	0.38328	0.8260
Blue-Yellow 1	0.56789	0.0920
Blue-Yellow 2	0.43212	0.6451

The separable nature of the kernels allows for the use of two relatively simple 1-D convolutions of the color planes, rather than a more complex 2-D convolution. The combination of positive and negative

weights in the achromatic channel creates a band-pass filter, as is traditionally associated with luminance contrast sensitivity functions. The positive weights used for the chrominance channels create two low-pass filters. Figure 18 illustrates the relative sensitivity of the three spatial filters, as a function of cycles per degree of visual angle, in both linear and log-log space. These plots were generated by performing a discrete Fourier transform (DFT) on the convolution kernels.

Figure 18. S-CIELAB Contrast Sensitivity Functions

The convolution kernels are used to spatially eliminate information in the frequencies that are imperceptible to the human visual system. The magnitudes of the remaining spatial frequencies are then normalized such that perceived color differences are the same for all frequencies. This is a similar concept to the Daly VDP in that it takes a larger magnitude difference in a higher spatial frequency to illicit the same color difference perception of a smaller magnitude difference at a lower spatial frequency.

The filtered opponent channels are then converted back from AC_1C_2 into $CIEXYZ$ tristimulus values using the equation below.

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} 0.979 & 1.189 & 1.232 \\ -1.535 & 0.764 & 1.163 \\ 0.445 & 0.135 & 2.079 \end{bmatrix} \begin{bmatrix} A \\ C_1 \\ C_2 \end{bmatrix} \quad (16)$$

The $CIEXYZ$ values for both the original and reproduction images are then converted into $CIELAB$ coordinates, using the white-point of the viewing conditions. A pixel-by-pixel color difference calculation can then be performed, resulting in an error image. Each pixel in the error image corresponds to the perceived color difference at that pixel. The spatial filtering assures that the color difference at each pixel is normalized to the traditional $CIELAB$ viewing conditions (simple patches) e.g. it takes a larger color difference at high frequencies to correspond to the color difference of simple patches. If desired, the error image can be converted into a single number that corresponds to the perceived image difference between

the image pairs. This can be accomplished using statistical methods, such as mean, median, maximum or RMS of the error images.

4.2 Color Visual Difference Model (CVDM)

The Color Visual Difference Model extends upon the idea of S-CIELAB. It too is a preprocessing step built onto traditional CIE color difference equations. The CVDM can actually be thought of as a hybrid between the S-CIELAB and VDP models discussed above. The general flowchart for the CVDM is shown below.³⁴

Figure 19. Flowchart of the CVDM, reprinted from Jin.³⁴

The model follows the same original path as S-CIELAB. The input images are first converted into the AC_1C_2 opponent space. This is followed by the spatial filtering stage. The spatial filtering can be performed using the same convolution kernels as S-CIELAB, if desired. Alternatively, the filtering can be performed in the frequency domain. The luminance channel is filtered using the CSF from the Visible Differences Predictor¹⁹, while the chrominance channels are filtered using simple low-pass filters.³⁴

The spatial frequency modulation is then followed by a Cortex transform.²⁴ This is identical to that used in the VDP for the luminance channel, and uses slightly fewer bands in the chrominance channels. A masking factor is calculated using the techniques described by Daly.¹⁹ This masking factor is used to calculate a visible difference factor using techniques similar to probability summation. This visible difference is then combined with the CSF filtered version of the original image (Image 1 in Figure

19) to calculate a new reproduction image. This is an important point, since the new reproduction image is essentially the original with only the visible differences added back to it. This step is necessary in order to calculate color differences, as the visible differences might include both positive and negative differences. The next stage is to convert the images are into CIEXYZ tristimulus values, and then into CIELAB coordinates. CIELAB is undefined for negative XYZ values, which is why the calculated differences are added back to the original image to create all positive XYZ units. A pixel-by-pixel color difference calculation is then performed to obtain an error image similar to S-CIELAB.

4.3 Magnitude Model Summary

Both S-CIELAB and the CVDM represent spatial-vision based models that are extensions to traditional CIE colorimetry. The S-CIELAB model is a very simple spatial pre-processing extension to CIELAB that has shown to be effective for predicting several psychophysical experiments.^{37,38} The relative success of this model indicates the potential of using traditional CIE color difference along with spatial pre-processing. The S-CIELAB model is hindered by its own simplicity, however.

The CVDM takes this idea one step further by combining the S-CIELAB model with the Visual Differences Predictor (VDP). In addition to spatial filtering, the CVDM adds orientation filtering and visual masking. All this is performed as pre-processing to the tradition color difference equations. The CVDM presents an interesting path for calculating magnitude color difference. As it stands it suffers from some of the same problems as the VDP, in that its complete implementation leaves several free parameters that need to be optimized to get accurate results for a given situation. The Cortex transform represents a rather costly computation as well, as it requires a pair of Fourier transforms for each sub-band. For the suggested number of sub-bands the model require at least 86 Fourier transforms to calculate a single image difference map.³⁴ The Laplacian pyramid along with the steerable orientation filters from the Sarnoff JND model could be used to create a similar multi-scale representation.

The idea of building a magnitude color image difference metric on top of traditional CIE color difference equations does look promising. A significant portion of the current research has focused on the creation of similar metrics. However, it is also of interest to examine more complicated, and complete, models of the human visual system.

5 Device-Independent Image Quality Modeling: Complex Vision Models

All of the perceptual models described so far are simple approximations of the human visual system, or loosely based on properties of the visual system. None of the models attempts to model the actual physiology of the human visual system. Such a process would be very complicated. Several existing models could be considered more complete representations of the entire visual system. These models were not designed specifically as image quality models, but rather overall models of low-level human perception.

Because these models were designed to be comprehensive spatial and color vision models, they can be used as part of a device-independent image quality model. While based on the physiology of the human visual system, they do not attempt to accurately model biological responses. Instead they behave somewhat as an empirical model or black-box. Often they are capable of predicting a wide range of spatial and color phenomena. Two such models are the Multiscale Observer Model^{39,40} (MOM) and the spatial ATD (Achromatic Tritanopic Deuternopic) model.⁴¹

5.1 *Multiscale Observer Model (MOM)*

The Multiscale Observer Model is designed to be a complete model of spatial vision and color appearance. It is capable of predicting a wide range of visual phenomena, including high-dynamic range tone-mapping, chromatic adaptation, luminance adaptation, spreading, and crispening.^{39,40} The general flowchart of the MOM is shown below. The flowchart should be considered an iconic representation of the complexity of the model, rather than an implementation guideline.

Figure 20. Flowchart of MOM, from Pattanaik.^{39,40}

The model appears rather complex, but it is actually similar in nature to many of the models already discussed. The first step is to take an input image, ideally a full spectral input image, otherwise a device dependent input image space such as CIEXYZ or LMS. This image is converted into fundamental cone signal images, LMS, plus a rod contribution image. These images are converted into 7 band-pass contrast images using the Laplacian pyramid technique described in the Sarnoff JND model.²⁰ Essentially the contrast image is a difference between a given band and another band of lower resolution. The band-pass images go through a gain modulation. This allows for flexibility in high dynamic ranges, as well as local adaptation. The adapted contrast signals are then converted into the opponent color space AC_1C_2 . This is the same color space used in S-CIELAB and the CVDM. The opponent color space representations are combined with the rod signal and then thresholded using nonlinear transducers. The result is a series of perceived contrast images. These images are converted back into cone signals, and then the Laplacian pyramid is collapsed back into single LMS bands. The LMS images can then be used as input into a traditional color appearance space to account for such things as chromatic adaptation.⁴²

It should be noted that the input to the MOM is defined as a single image. This model was designed to predict the appearance of an image based solely upon the information contained in itself, and the viewing conditions. This can be extended into image difference or image quality by processing a second image, and then comparing the appearance of, or the difference between, the two images.

The Multiscale Observer Model is an example of a very comprehensive model of both spatial and color appearance. This type of model could be capable of augmenting or replacing psychophysical experimentation in the device-dependent approach to image quality. In order to be used in this fashion it would be necessary to implement the model so that it predicts certain variables of perception, such as sharpness, graininess, and contrast. These predictions could then be verified against psychophysical data. Likewise, adding a second image allows this type of model to be used in a device-independent approach to image quality. The ability to actually predict appearance correlates as well as image differences is another strength of this model. Those correlates could then be used to better predict the perception of quality rather than just image differences. It should be noted that currently these types of analyses have not been performed with this model, due in part to the inherent difficulty in implementation. The weakness of the MOM lies mostly in its complexity. It is a difficult model to implement, and a computationally expensive model. There are also a number of free parameters that need to be adjusted depending on the application. These parameters can be calibrated and fit to experimental data sets, if they exist. The MOM also has no provisions for orientation filtering, though that could be easily remedied using the steerable transform filters, which is a similar approach to the Sarnoff JND metric.²⁰

5.2 Spatial ATD

The spatial ATD model is a modification of the original ATD model of color perception and visual adaptation published by Guth.⁴³ The model was adapted by Granger⁴¹ to include a model for spatial frequency filtering. This adaptation was similar in nature to the spatial filtering extension in S-CIELAB. The general flowchart for the Spatial ATD model is shown below in Figure 21.

Figure 21. Flowchart for the Spatial ATD Model.

The input into the model is an RGB image. This image is first converted into cone responses using the characterization of the viewing conditions. One major difference between the ATD model and all the previously described models is that the cone signals themselves are non-linear transformations of CIE XYZ tristimulus values with an additive noise factor. The additive noise is an empirically derived constant that varies for each of the LMS cone responses. The nonlinear cone signals are then processed

through a gain-control mechanism, which accounts for chromatic adaptation. The adapted signals are then transformed into an opponent color space, ATD for achromatic, tritanopic, and deuteranopic. The opponent color space coordinates are then filtered using a band-pass filter for the luminance channel (A) and two low-pass filters for the chrominance channels (TD). As published, there are no specifics for the components of the spatial filters, other than their general shape. The spatially filtered signals are then subjected to a nonlinear compression function, which accounts for luminance and potentially chromatic adaptation. Finally, the compressed ATD signals could be transformed into appearance correlates, although the method for calculation is also not specifically defined.

If an image difference is desired, two images can be processed simultaneously. The color difference can then be taken directly from the compressed ATD coordinates, or from the appearance correlates. There has been little research in developing a precise technique for calculating color difference using this color space.

The ATD model itself has been slightly altered since the original inception of the spatial extension.⁴³ This alteration did not change the general form of the model, and thus the spatial filtering should still be applicable. The spatial ATD model is somewhat of a hybrid between the vision-based magnitude models and the more complete Multiscale Observer Model. It presents a simple model of the human visual system that can be capable of predicting psychophysical data.⁴¹ The relative simplicity of the model does not allow it to predict more complicated spatial and color appearance phenomena. The ATD model itself also suffers from a lack of clear definition, as there are many free parameters that need to be better defined in order to be considered a full model of color and spatial appearance.⁴²

5.3 Summary of Complex Visual Models

The complex visual models described above strive to be complete models of spatial and color vision. While not strictly following the actual physiology of the human visual system, they are empirical models that behave similarly to experimental results. Both the MOM and the spatial ATD model are capable of predicting both image color differences and appearance attributes. These appearance attributes should allow for an easier correlation between perceived differences and image quality. The fundamental ideas stressed in these complex vision models will be revisited in later sections. Neither model has a documented technique for predicting color differences, so it is unknown how these models relate to traditional color spaces.

6 General Framework for a Color Image Difference Metric

Thus far we have reviewed many historical approaches to image quality modeling. These approaches vary in technique as well as general goals. The device-dependent approaches to image quality modeling attempt to directly link imaging system parameters with human perceptions. The device-independent approaches attempt to relate properties of the images themselves with human perception. Although the approaches might differ, the ultimate goal of any image quality model is to mathematically predict perception. Much can be learned by examining the historical approaches to image quality.

The sheer number of researchers, along with the number of different approaches, indicates that image quality modeling is a very complicated task. Device-dependent approaches have proven to be successful in the design and evaluation of complete imaging systems.^{5,6,7,44,45,15} These approaches typically require exhaustive psychophysical evaluation to correlate system variables with perceived image quality. Several perceptually based models have been designed to potentially eliminate the need for the psychophysics, or at least augment the experimental design.

These perceptual models can be separated into two distinct categories: threshold and magnitude models. The threshold models are excellent at predicting whether an observer will perceive a difference between two images. The magnitude models strive to predict the size of the perceived difference. What follows is a generalized approach to the formation of a new perception-based magnitude difference model designed to build upon the strengths of all the models described above, while eliminating several weaknesses.

The first goal of this research project is the formulation of a general framework for the creation of an image difference metric. This framework is designed with three concepts in mind:

- *Simplicity*
- *Use of Existing Color Difference Research*
- *Modularity*

Recall the automobile analogy discussed in Section 1.3.1, the framework can be thought of as the general shape of the “car.” The modular nature allows for building of the image difference module from various “off the shelf” components.

6.1 Framework Concept: Model Simplicity

The color image difference metric should be as simple as possible. This seems like an obvious goal, but in practice is much more difficult than it sounds. The framework for the development of a metric should emphasize techniques that are relatively simple in implementation and concept. This does not

imply that the framework for model development should not allow for any complex calculations, but rather that each calculation is well designed and understood.

If a model is simple to implement then it has a much greater chance of reaching a widespread audience. If many researchers can implement and test a model, then many researchers can also contribute to the growth and improvement of the model. This is very beneficial because the more testing any model receives, the more accurate it will eventually be. One only needs to look at the complexity of the Multiscale Observer Model to understand why that model has not been adopted universally by both researchers and industry. Another complexity that should be avoided, as illustrated in many of the models described in the previous sections, is the use of free parameters that require fitting in order to be used. Free parameters can be allowed, and are often beneficial, as long as clear usage guidelines are also available.

Simplicity as a generalized concept for the model also allows for a much greater understanding of each stage in the calculation. With this as a goal, it becomes possible to test, and potentially improve upon, every element of the model. If each of the elements in a model interact at various stages of calculations, it becomes increasingly difficult to understand the importance of any given element. This concept is similar in nature to the Modularity goal of the framework.

Another concept that might fall under the umbrella of simplicity is the idea of computational cost. If an image difference metric takes too long to calculate, then the benefit of the metric is reduced. While this problem will inevitably fade as computers increase in speed, it is still a reality on current hardware. The computational cost must also be balanced against the cost of performing the psychophysical experimentation the model is designed to augment. If it is easier and cheaper to perform actual experiments then the benefits of computational models disappear.

6.2 Framework Concept: Use of Existing Color Difference Research

Equations and models for specifying color difference have been a topic of study for many years. This research has culminated in the CIE DE94/2000 color difference equations.^{46,47} These equations have proven to be successful in the prediction of color differences for simple color patches, as well as developing instrumental based color tolerances. Since they were derived using color patches in well-defined viewing conditions, their use in color imaging is less apparent. While these models were never designed for color imaging applications, the successes they enjoy as well as their adoption for industrial use, serve as a good foundation upon which to build.

This was the concept generalized by the S-CIELAB model.³³ The research framework presented here will follow a similar path. If the model collapses into existing color difference equations when presented with large uniform stimuli, then it is possible to have a single color difference metric that can

predict both spatially simple and complex stimuli. By focusing the framework to build upon existing color difference research, there is no need to “re-invent the wheel.”

6.3 Framework Concept: Modularity

Modularity is a very important design goal for the image difference framework. The idea of modularity is to allow the different functional elements in the eventual image difference model to be removed or replaced, much like building blocks. Self-contained “modules” assure that the removal of any single element in the model will not remove the functionality from any of the other. This allows for a general evolution of the final color image difference metric. Modularity also ties heavily with the goal of simplicity. With a modular framework, we can first choose a relatively simple core metric, such as the CIE color difference equations, and then build calculations that are more complicated on top of that, as they are deemed necessary. If the simple model is accurate enough, then there is no need for the more complicated “modules.” If the need for more complexity arises, then other modules can be designed and utilized. Both S-CIELAB and the CVDM are examples of this type of building system. S-CIELAB is an added spatial filtering module on top of CIELAB. The CVDM goes another step further and adds a module on top of S-CIELAB to predict visual masking.

The development of a modular image difference framework might take two potential approaches. The first concept follows the hierarchy of most of the models described in the previous sections. This is the concept of the “building block” technique, where each module in the framework builds upon the other. This concept is illustrated below in Figure 22.

At the base of the “structure” is the core metric, such as the CIE color difference equations. Each element is then built upon the previous module or computation. This type of framework requires a strict order of the elements. If the modules are not self sufficient, meaning they require other blocks in order to function, then some of the modularity in the system is lost. Care should be taken to assure that there are not too many interdependencies between the modules. Generally, for this type of framework, the order in which the blocks are stacked is very important. The building-block technique should not eliminate the potential for each block to be removed, or replaced. Even if there are interdependencies, the blocks should be able to evolve, and be replaced as experimental testing warrants.

Figure 22. Concept of a Modular Building Block Framework

Another general framework concept can be considered more of a freeform pool of modules. There still is a core metric at the heart of this type of framework, upon which the full structure of the model is built. In this type of design the structure is not as rigid as it is in the building block framework. This concept is illustrated below in Figure 23.

Figure 23. General Concept of a Modular Pool Framework

In this type of framework there is a core metric, and then a “pool” of available modules. Depending on the application, the user can select any of the available modules to combine with the core metric. In the above figure, if users are concerned with detecting changes in image contrast they might choose to use only the local contrast metric while ignoring all other modules. If they are interested in image sharpness, they might choose the attention module as well as the local contrast module. The general idea is that each module is a self-contained unit. As such, each receives both input and has output. Often the output is fed directly into another module, and eventually into the core metric. This does not have to be the case, however. Each module can be designed to maintain its own output, to be pooled later into a more generalized model. For instance, it should be possible to determine if there is a magnitude difference between images using the core metric, and then to determine if that difference was a result of a contrast change by examining the output of a “contrast module.”

It is important to note that the concepts illustrated in Figure 22 and Figure 23 are not mutually exclusive. In practice, they often reduce to the same framework. For this research project, both concepts were considered. The framework for the image difference metric allows for a “pool” of modules from which the user can select. Once the modules are selected, the order in which they are applied becomes

important. Thus, the selected modules are then placed into an ordered building block structure. If the order of application were not important, then we would have the most flexible modular framework.

6.4 Framework Evaluation: Psychophysical Verification

It is important to gather psychophysical data to both develop and evaluate the various modules in the image difference framework. These data could also be used to determine the necessary order of the various modules, though in most module the order of application is implied and should not be altered to fit empirical data. Several psychophysical experiments are described in later sections that were used for model development and testing. These experiments were designed to test certain individual module design parameters as well as the overall image difference metric.

As described above, it is important that the image difference metrics are both simple and flexible. With this in mind, the goal is not to strictly fit empirical equations to large amounts of psychophysical data. This might prove to be a successful device-dependent modeling approach. The goal is not to characterize the image quality of any given imaging system, so instead this research will focus on a device-independent approach to quality modeling. Each of the modules is designed using a theoretical approach taking cues from the perception-based modeling done in the VDP and the SQRI. These theoretical models are then tested and fit against the psychophysical data.

While the research goal is not in the creation a strictly empirical framework, it is possible to design psychophysical experiments that can be used to tune certain modules or general aspects of the framework in general. When models are fit to the experimental results, it becomes important to test those fits against other independent data in order to validate them. Several independent experiments that were used to validate the modules are described in Section 8.

6.5 General Framework: Conclusion

This section outlined a general framework for developing a color image difference metric. This framework represents a step towards the first goal of this research project, which is the creation and evaluation of a perception-based image difference model capable of predicting magnitude differences. Three main concepts of the framework were discussed: simplicity, modularity, and the use of existing color difference equations. The framework itself is an important step towards the research goals because it allows for great flexibility for model development. The next section outlines several specific modules that have been designed to fit into this framework.

7 Modules for Image Difference Framework

The previous section outlined the general framework that guided the development of a color image difference metric. This section introduces several individual modules of that framework, which can be used to build a comprehensive image difference metric. The majority of these modules are inspired by the S-CIELAB spatial filter pre-processing to the CIELAB color space. As such, most of the modules described below use CIELAB, and specifically the CIE color difference equations as the “core metric.” Perhaps appropriately, the first module discussed is spatial filtering using the contrast sensitivity function of the human visual system. Other modules discussed include spatial frequency adaptation, a local and global contrast metric, a type of local attention metric, and error summation and reduction.

The final module discussed is the actual color space used for the difference calculations. This can be thought of as the module for the “core metric.” The color space does not necessarily have to be CIELAB, though that does have the benefit of many years of color difference research. It might be more appropriate to use appearance spaces, such as CIECAM02, if trying to measure changes in appearance between two images across disparate viewing conditions.

Many of the modules described below have been discussed in detail in publications.^{48,49}

7.1 Spatial Filtering Module

The first module discussed is inspired by the S-CIELAB spatial extension to traditional CIELAB.³³ S-CIELAB was described in greater detail in Section 4.1 above, and is graphically represented in Figure 16. Essentially, the S-CIELAB model uses CIE color difference equations such as $CIE \Delta E_{ab}^*$ and $CIE \Delta E_{94}$ in conjunction with spatial filtering. The spatial filtering is performed as a pre-processing step, and is used to approximate the spatial properties of the human visual system. In the context of the modular image difference framework, S-CIELAB uses CIELAB as a core metric, and adds to it a spatial filtering module.

The spatial filtering in S-CIELAB is performed using a series of 1-D separable convolution kernels on an opponent color space. These kernels are designed to approximate the contrast sensitivity functions of the human visual system. The CSF is often used to modulate spatial frequencies that are less perceptible to a human observer. For this reason, the CSF is often erroneously referred to as the modulation transfer function (MTF) of the human visual system. While similar in nature to an MTF, specification of a CSF makes no implicit assumption that the human visual system behaves as a linear system.¹⁹ Rather it is better to take a similar thought process as Barten¹⁰ and Daly.¹⁹ That is to think of the CSF as a way of normalizing spatial frequencies such that they have equal contrast thresholds. In the case of magnitude image differences, this implies that the color difference is normalized for all frequencies.

Fourier theory dictates that the discrete convolution kernels allow only for the sum or difference of cosine waves. These cosine waves are in effect only an approximation of more accurately defined contrast sensitivity functions, when used with kernels of limited size. As the size of the convolution kernel increases the convolution kernel becomes identical to the frequency filter. This approximation is balanced out by the ease of implementation and computation of the convolution. Specifying and implementing the contrast sensitivity filters purely in the frequency, rather than spatial domain, allows for more precise control over the filters with a smaller number of model parameters. Spatial filtering in the frequency domain follows the general form shown below:

$$Image_{filt} = \mathcal{F}^{-1} \{ \mathcal{F} \{ Image \} \cdot Filter \} \quad (17)$$

where $Image_{filt}$ is the filtered image, $\mathcal{F}^{-1}$ and $\mathcal{F}$ are the inverse, and forward Fourier transform respectively, $Image$ is the original input image, and $Filter$ is the 2-D frequency filter.

The convolution kernels from the S-CIELAB model create a band-pass filter for the luminance opponent channel, and two low-pass filters for the red-green and yellow-blue chrominance channels. It would be possible to simply use the Fourier transform of these kernels, as shown in Figure 18. This does not gain any benefit over the convolution approach. Rather new filters can be designed that are potentially more precisely related to measured contrast sensitivity functions than the S-CIELAB approximations.

The chrominance filters from the S-CIELAB model were fit to experimental data collected by Poirson and Wandell.³⁵ Those data can be combined with other experimental data such as that from Mullen⁵⁰, or Van der Horst and Bouman⁵¹. The sum of two Gaussian functions were fit to the Poirson and Van der Horst data using non-linear optimization. The fit of the Gaussians was very good for the independent data-sets, as well as the combined sets. The form of the Gaussian equations are shown below:

$$CSF_{chrom}(f) = a_1 \cdot e^{-b_1 \cdot f^{\alpha_1}} + a_2 \cdot e^{-b_2 \cdot f^{\alpha_2}} \quad (18)$$

where f is spatial frequency in cycles per degree of visual angle. Table 3 shows the values of the six parameters for the red-green, and blue-yellow equations that best fit the combined data sets. Figure 24 shows the normalized sensitivities of the two chrominance channels, as a function of cycles per degree of visual angle. It is important to note that these filters are isotropic in nature, as there is not enough oblique data available to fit anisotropic chrominance filters.

Table 3. Parameters for Chrominance CSFs

Parameter	Red-Green	Blue-Yellow
$a1$	109.1413	7.0328
$b1$	-0.0004	0.0000
$c1$	3.4244	4.2582
$a2$	93.5971	40.6910
$b2$	-0.0037	-0.1039
$c2$	2.1677	1.6487

Figure 24. Red-Green and Yellow-Blue Frequency CSF Filters.

As these filters are to be applied to the 2-D frequency representation of the Red-Green and Blue-Yellow channels, they must also be 2-D filters. The 2-D representation of these filters is shown in Figure 25.

Figure 25. 2-D Representation of Red-Green (left) and Yellow-Blue (right) Filters

The luminance filter should be a band-pass filter, to approximate the contrast sensitivity function of the human visual system. The Fourier transform of the S-CIELAB convolution kernels was shown in Figure 18. These filters were designed to fit experimental data. A three parameter exponential equation, described by Movshon⁵², is a simple description of the general shape of the luminance CSF, which behaves similarly to the S-CIELAB filter. The form of this model is shown below:

$$csf_{lum}(f) = a \cdot f^c \cdot e^{-b \cdot f} \quad (19)$$

where values of 75, 0.2, and 0.8 for a , b , and c respectively approximate a typical observer.⁵³ The general shape of this function is shown in Figure 26. This model requires only three parameters, as compared to the many parameters of the Daly and Barten CSF models. Its simplicity comes as a cost though, as it does not change as a function of viewing conditions.

Figure 26. General Shape of Three Parameter Movshon CSF

This filter will also be applied to a two-dimensional image, and as such must be two-dimensional. The two-dimensional representation is illustrated below.

Figure 27. Two-Dimensional Representation of Movshon CSF

It is important to note that the above filter behaves as a band-pass filter, peaking around 4 cycles-per-degree. Careful consideration needs to be taken regarding the DC component of the filter. The DC component is essentially the mean value of the image channel. For large simple patches, the mean value is the value of the patch. The existing color difference formulas are able to accurately predict color differences of simple patches, so it is important to keep this mean value constant. This can be accomplished in several ways. Either the luminance contrast sensitivity function can be truncated into a low-pass filter, or it can be normalized such that the DC component is equal to unity. If the latter method is chosen, then the spatial filtering behaves as a frequency enhancer as well as modulator. Examples of both these DC frequency-maintaining techniques are illustrated in Figure 28.

Figure 28. Examples of DC Maintaining Luminance CSF using Low-pass (dashed line) or Frequency Enhancing (solid line)

The relative sensitivities of the band-pass filter in Figure 28 include values that are greater than 1.0, and peak around 4 cycles per degree of visual angle. Instead of simply modulating frequencies, this actually serves to enhance any image differences where the human visual system is most sensitive to them. When attempting to predict the perceived visual differences between two images this frequency enhancement boosts color differences at certain frequencies and modulates (or eliminates) color differences at other frequencies.

The relatively simple form of the Movshon three-parameter equation is both its strength and weakness. It is an isotropic function so it is incapable of predicting orientation phenomena such as the oblique effect. The form of the function is the same for all viewing conditions, unless the parameters are specifically fit to existing data. It is generally assumed that viewing conditions can greatly alter the contrast sensitivity function. This is especially the case for luminance level, which is known to “flatten” the shape of the contrast sensitivity function as luminance increases. To better predict changes in viewing condition, a more complicated function might be necessary. Two such functions have already been discussed in the form of the SQRI and VDP models from Sections 2.3 and 3.1 above.

7.1.1 Barten CSF from Square Root Integral Model (SQRI)

The contrast sensitivity function described by Barten for the SQRI model was described in detail in Section 2.3. The contrast sensitivity model begins with the optical MTF of the human eye, which is expressed as a Gaussian function. The MTF is then modified with models of photon and neural noise, and lateral inhibition. The resulting CSF is an isotropic band-pass shape that is a function of luminance level,

pupil diameter, and image size. The general shape of this function was shown in Figure 7. The Barten CSF has the same general band-pass shape as the Movshon CSF, resulting in modulation of the DC component. This can be taken care of using the same normalization techniques as described above, resulting in a CSF that both modulates and enhances. The two-dimensional CSF functions for use in the image difference framework presented here are shown below.

Figure 29. Two-Dimensional Representation of Barten Luminance CSF

Since the Barten CSF is a function of many viewing condition parameters it is inherently more flexible than the Movshon CSF. This flexibility comes with the price of model complication, as shown in Equation 7. This model also predicts an isotropic contrast sensitivity function, so it too is incapable of predicting orientation effects.

7.1.2 Daly CSF from the Visual Differences Predictor (VDP)

The contrast sensitivity function from the VDP was also described in detail in Section 3.1. The general form of this model is shown in Figure 9. This model is a function of many parameters, including radial spatial frequency (orientation), luminance levels, image size, image eccentricity, and viewing distance. The result is an anisotropic band-pass function that represents greater sensitivity to horizontal and vertical spatial frequencies than to diagonal frequencies. This corresponds well with the known behavior of the human visual system (the oblique effect). The two-dimensional representation was shown in Figure 10 and Figure 30.

Figure 30. Two-Dimensional Daly CSF Filter

The Daly CSF provides a comprehensive model capable of taking into account a wide range of viewing conditions. The weakness in this model is in its complexity.

7.1.3 *Modified Movshon*

Another potential approach is to modify the three-parameter Movshon model such that it can handle a wider range in viewing conditions. By altering the three parameters a , b , and c as a function of adapting luminance it would be easy to add luminance factors to the model. Similarly, it would be possible to combine the orientation function from the Daly model with the simple form of the Movshon model. The form of this is shown below:

$$\text{csf}_{\text{lum}}(f, \theta) = a \cdot f_{\theta}^c \cdot e^{-b \cdot f_{\theta}} \quad f_{\theta} = \frac{f}{\frac{1-r}{2} \cos(4\theta) + \frac{1+r}{2}} \quad (20)$$

Where r represents the degree of modulation desired for diagonal frequencies, which is 0.78 for the Daly model. This allows for a relatively simple model capable of predicting orientation effects. The simple model sacrifices some flexibility because it does not change with viewing conditions. The Daly and Barten models change their shape as functions of viewing conditions such as overall luminance level, viewing distance, and image size. The differences between these contrast sensitivity functions are discussed in further detail in Section 9.1.2.

7.1.4 *Spatial Filtering Summary*

The above sections describe a spatial filtering module for the image difference framework. The pre-processing filtering described in the S-CIELAB model serves as the inspiration for this module. The spatial filtering is based on the contrast sensitivity function of the human visual system. It is important to stress that the CSF does not describe an actual physiological mechanism. Rather it is a function used to describe the threshold detection behavior of the human visual system. The CSF is being used as frequency filters to approximate this behavior. The goal of filtering with the CSF is to normalize all spatial frequencies such that the magnitude for color difference is identical. This essentially means that it takes more of a physically measurable difference at high frequencies to illicit the same perceptual response as a smaller physically measured difference at a lower frequency. If the CSF function goes above 1.0 then it serves to also enhance some frequencies. For instance the perceived color difference of two stimuli at 4 cycles-per-degree would look larger than the same stimuli at higher or lower frequencies.

These contrast sensitivity functions of the human visual system are described as band-pass for the luminance channel, and low-pass for the chrominance channels. Whereas S-CIELAB approximated these functions using a series of one-dimensional convolution kernels, all of the models described above are specified as frequency-space filters. These filters could also be used as 2-D convolution kernels by taking the Discrete Fourier Transform (DFT) of the frequency filters, if desired. The general form of an image difference metric using CIELAB color difference equations and the spatial frequency module is illustrated in Figure 31.

Figure 31. Flowchart for Spatial Filtering Module

7.2 Spatial Frequency Adaptation

There are several techniques for measuring the contrast sensitivity function of a human observer. Most often it is through the use of a simple grating pattern.⁵⁴ This pattern is usually flashed temporally to prevent the observer from adapting to the spatial frequencies being tested. Spatial frequency adaptation, similar to chromatic adaptation, results in a decrease in sensitivity based on the adapting frequency.⁵⁵

While spatial-frequency adaptation is not desired when measuring the fundamental nature of the human visual system, it is a fact of life in real-world imaging situations. Models of spatial-frequency adaptation that alter the nature of the contrast sensitivity function might be better suited for use with complex image stimuli than those designed to predict simple gratings. The effect of frequency adaptation

is to boost and shift the peak of the CSF, when normalized with the DC component. Essentially this implies that we adapt more in the low frequency regions than the higher regions.⁵⁶ Figure 32 illustrates this concept. What this concept implies is that the peak of the contrast sensitivity function might perhaps be at higher frequencies with spatially complex stimuli, and that overall sensitivity might actually increase. This behavior has been noticed before in several experiments.^{56,57} Two such models of spatial frequency adaptation are presented in the following sections.

Figure 32. General Concept of Spatial Frequency Adaptation Illustrating Adaptation More to Lower Frequencies than Higher Frequencies

7.2.1 Natural Scene Assumption

It is often assumed that the power of any given frequency in the “natural world” is inversely proportional to the frequency itself.⁵⁶ This is known as the $1/f$ approximation. If this assumption is held to be true, then the contrast sensitivity function for natural scenes should be decreased more for the lower frequencies, and less for higher. When the CSF is renormalized so that the DC component is unity, the result is a CSF that illustrates an increase in relative sensitivity, with the peak shifted. It should be noted that this relative increase in sensitivity has actually shown up in experimental conditions using images as adapting stimuli, and then measuring CSFs using traditional sine-wave patterns.⁵⁶ Recall also the logarithmic integration utilized by the SQF and SQRI function was also in effect utilizing a $1/f$ modulation of the CSF. Equation 21 shows a simple linear “von Kries” type of adaptation based on this natural-world assumption:

$$csf_{adapt}(f) = \frac{csf(f)}{1/f} = f \cdot csf(f) \quad (21)$$

Where f represents the spatial frequency in cycles per degree. In practice this type of spatial adaptation modulates the low frequencies too much, and places too high an emphasis on the higher frequencies. The effect of this adaptation is illustrated in Figure 33.

Figure 33. Spatial Frequency Adaptation with 1/f

Figure 33 clearly shows the shift in the peak, as well as the boost of the CSF. The resulting boost is far too large as it now boosts certain frequencies by a magnitude of 15. A nonlinear compressive function is better suited for imaging applications. The form of this equation is shown below:

$$csf_{adapt}(f) = \frac{csf(f)}{\left(\frac{1}{f}\right)^{1/x}} = f^{1/x} \cdot csf(f) \quad (22)$$

where f is the spatial frequency and x is the compressive exponent. Figure 34 shows the resulting contrast sensitivity functions using compressive exponents of 2 and 3 (for exponents of 1/2 and 1/3). The frequency enhancements shown in Figure 34 are much less than those without the compressive exponent. The peak for the 1/2 compressive function roughly 4 units at 9 cycles-per-degree, while the peak of the 1/3 is 2.8 units at 7 cycles-per-degree. The shape of the CSF is also important for imaging applications. If the peak is too sharp (too tall and narrow) then the artifact of ringing can occur. Figure 35 illustrates the artifact of ringing when an image is filtered with a “sharp” CSF.

Figure 34. Spatial Frequency Adaptation with a Compressed 1/f

Original

Filtered with "Sharp CSF"

Figure 35. Ringing Artifacts Caused by Filtering with "Sharp" CSF

The balance between adapting the CSF in a meaningful way and avoiding ringing artifacts is balanced with a compressive exponent of 1/3. This serves to boost the entire CSF and shift the peak to 7 cpd while not causing any ringing artifacts.

7.2.2 *Image Dependent Spatial Frequency Adaptation*

A more complicated approach to spatial frequency adaptation involves adapting to the frequencies present in the image itself, rather than making assumptions about the frequency power present in the natural world. In this case, the contrast sensitivity is modulated by the percentage of occurrence of any given frequency in the image itself. The frequency of occurrence can be thought of as a frequency histogram. The general form of this type of adaptation is shown below:

$$csf_{adapt}(f) = \frac{csf(f)}{histogram(f)} \quad (23)$$

where f is the spatial frequency in cycles per degree of visual angle, and $histogram(f)$ represents the frequency of occurrence of any given spatial frequency in the image. The frequency histogram can be obtained by taking the Fourier transform of the image. Again, this idea is more difficult in practice. For most images, the DC component represents the majority of the frequencies present. This tends to overwhelm all other frequencies, and results in complete modulation of the DC component, and gross exaggeration of the very high frequencies. This turns the CSF into a highpass filter, which is the exact opposite of what is desired. One way around this problem is to clip the percentage of the DC component to a maximum contribution. We have found that 10% represents a reasonable DC contribution for most images. This is illustrated in the Equation below.

$$histogram(f) \approx fft(image) < 10\% \quad (24)$$

The resulting function is still incredibly noisy, and prone to error as some frequencies have very small contributions, while others have much larger contributions. These small contributions correspond to very large increases in contrast sensitivity when normalized using Equation 23. This problem can be eliminated by smoothing the entire range of frequencies using a statistical filter, such as a Lee filter.⁵⁸ These filters compute statistical expected values based on a local neighborhood, and are specifically designed to eliminate noise. The resulting normalization still places too large an emphasis on the high frequencies as compared to the lower frequencies. This can be overcome by applying the same nonlinear compression function that was described for the natural scene assumption. Care must be taken to avoid ringing from CSFs that are too sharp. The final form of an image dependent spatial frequency adaptation equation that balances these criteria is shown below.

$$csf_{adapt}(f) = \frac{csf(f)}{\left(smooth[fft(image) < 10\%] \right)^{1/3}} \quad (25)$$

7.2.3 Spatial Frequency Adaptation Summary

Two models of spatial frequency adaptation are described above. These models serve to alter the CSF functions described in the previous module discussion. As such, these models cannot be thought of as independent modules in the framework presented. Rather they are cascaded along with the general spatial filtering module. Thus the general flowchart for applying these functions is the same as shown in Figure 31. The following sections outline some modules that can be used independently or in conjunction with the spatial filtering module.

7.3 Spatial Localization Filtering

The contrast sensitivity filters as described above generally serve to decrease the perceived differences for high frequency image information, such as halftone dots. However, it is often observed that the human visual system is especially sensitive to position of edges. The contrast sensitivity functions seem to counter this theory, as edges contain very high frequencies. This contradiction can be resolved if we consider this a type of localization.

The ability to distinguish, or localize, edges and lines beyond the resolution of the cone distribution itself is well documented.^{54,59} While the actual mechanisms of the human visual system might not be known, it is possible to create a simple module to account for this ability to detect edges.

7.3.1 Spatial Localization: Simple Image Processing Approach

The simplest such approach to this type of modeling is borrowed from the image-processing world. Edge detection algorithms are very common in image processing, and can be easily applied in the context of an image difference model. An example of this type of processing is convolution with a two-dimensional Sobel kernel. The general form of this kernel is as follows.

$$x_{dir} = \begin{bmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{bmatrix} \quad y_{dir} = \begin{bmatrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ 1 & 2 & 1 \end{bmatrix} \quad (26)$$

The benefit to this type of localization kernel lies in its simplicity, and the fact that most computing languages have pre-defined libraries for implementing such a kernel. The drawback to this filter is that it does not take into account the viewing distance unless the images are preprocessed to a set number of pixels-per-degree, so as a result is not as well tuned for all applications. The frequencies being enhanced will change as a direct result of the viewing condition. In order for the image difference framework to be flexible with regards to viewing conditions, it would be a wise idea to have a localization filter that is more flexible.

7.3.2 Spatial Localization: Difference of Gaussian

A more flexible type approach to a spatial localization kernel would be to use a tunable Difference-of-Gaussian (DOG) Filter. The general form of this type of equation is as follows:

$$DOG = e^{-\left(\frac{x^2+y^2}{\sigma_1^2}\right)} - e^{-\left(\frac{x^2+y^2}{\sigma_2^2}\right)} \quad (27)$$

where x and y are the two-dimensional spatial image coordinates (pixels), and σ_1 and σ_2 are the relative widths of the two Gaussian functions. The general shape of this type function is illustrated in Figure 36.

Figure 36. General Shape of Difference-of-Gaussian Filter

This type of kernel can be frequency tuned by altering the values of σ_1 and σ_2 to the desired widths. Once the desired widths are chosen, the DOG kernel is applied to the image using a standard two-dimensional convolution process. The DOG kernel is popular in part because of the similarities between its shape and aspects of human physiology. The human visual system is composed of center surround ganglion cells that have similar shape to the difference of Gaussian kernel.⁶⁰ These center surround cells can have an excitatory center and inhibition surround (called on-center) or vice-versa (off-center). It should be stressed that convolution with an on-center DOG is equivalent to frequency filtering with a band-pass filter. Depending on the size of the two Gaussians this filter can be tuned to specific peak frequencies and bandwidth.

7.3.3 *Spatial Localization: Frequency filtering*

This same type of localization filtering can be accomplished by multiplication in the frequency domain, rather than convolution in the spatial domain. This approach has the benefit of being able to more intuitively select the frequencies that should be enhanced. For instance, it is easy to specify a band-pass Gaussian filter centered at 30 cycles-per-degree with a half-width of 5 cycles-per degree. An example of this type of filter is shown in Figure 37

Figure 37. Example of a Band-Pass Spatial Localization Filter

The process of applying this filter would be the same as that for applying the contrast sensitivity functions. The image is first transformed in the frequency domain via the Fourier transform, and then multiplied by the spatial localization filter. The result is then processed through an inverse Fourier transform to obtain the filtered image. If the frequency filter shown in Figure 37 is converted into the spatial domain it will look similar to the DOG kernel.

Since the process of frequency filtering is identical to that of the CSF filtering it stands that the two processes could be cascaded together into a single filter. This process has both benefits and drawbacks. The benefit is that the CSF filtering and spatial localization filtering can occur simultaneously. An example of this type of calculation would be to first multiply the CSF by the localization filter. This multiplication is shown in Figure 38. The drawbacks to this cascading approach lie in the sacrificing of modularity for the features. It becomes more difficult to separate the CSF from the local attention filter for future testing and enhancements. For many cases this lack of modularity is not a problem, as there can be a single module for “spatial filtering” that can envelope all of the various spatial filters. The benefits of this single filtering module that can contain both the CSF modulation and spatial localization filtering, along with any filters designed in the future, is clearly evident.

Figure 38. Example of Cascading CSF and Spatial Localization Filter into Single Filter

7.3.4 *Spatial Localization: Summary*

Three techniques for performing spatial localization filtering were described. Spatial localization attempts to model the human visual systems ability to detect edge information, often beyond the resolution of the cone spacing itself. The spatial localization module can be applied to an image difference model as a stand-alone module in conjunction with the core metric, or it can be cascaded with the spatial frequency-filtering module.

One possible benefit to using the localization module on its own lies in the creation of a “sharpness” map. While the output of the module should be fed into the core metric for calculating color differences, the sharpness map might provide insight into whether the perceived differences are caused by changes in perceived sharpness. Further research is needed to specifically create a link between the localization output, and actual perceived sharpness.

7.4 **Local and Global Contrast**

The ability of an image difference model to predict both local and global perceived contrast differences is very important.³⁸ This can be considered another area where localization and attention play a factor. Image contrast can often be thought about in terms of image tone reproduction. Moroney⁶¹ presented a local color correction technique based on non-linear masking, which essentially provided a local tone reproduction curve for every pixel in an image. This technique, which was designed to enhance local contrast, can be adapted to provide a method for detection and prediction of image contrast differences.

This color correction technique generates a family of gamma-correction curves based upon the value of a low-frequency image mask. This can be extended to an image difference model by generating a family of gamma curves for each pixel in the image, based not only on the low frequency information at each channel, but also the global contrast of each channel. The low-frequency mask for each image can be generated by filtering each image with a low-pass Gaussian curve. An example of a low-pass mask is shown below in Figure 39. It is often helpful to use a modified Hanning window to reduce ringing artifacts in the mask.

Figure 39. Example of an Original Image and its Low-pass Mask

The contrast curves can then be generated using a modified form of Moroney's technique, while accounting for the use of a positive image mask. The general form of this equation is as follows:

$$gamma[x,y] = \max \cdot \left[\frac{image[x,y]}{\max} \right]^2 \left\{ \frac{median}{median-mask[x,y]} \right\} \quad (28)$$

where $gamma[x,y]$ is the tone reproduction curve generated for each pixel, $image$ is the input image channel, $\max$ is the maximum value of the image in that channel, $median$ is the median value of the image in that channel, and $mask[x,y]$ is the value of the low-pass mask for a given pixel location. The use of the image channel maximum and medium helps assure that images of different "global" contrast values will create different families of tone reproduction curves. This should serve to predict global changes in contrast between two images. The family of curves generated using Equation 28 are shown in Figure 40.

Figure 40. Family of Tone Reproduction Curves Generated Using Local Contrast Model

7.4.1 Local and Global Contrast Summary

This rather simple local and global contrast detection model has tremendous potential and flexibility. It can be used in the image color difference framework as an independent module, or it can be used in combination with other modules to create a more complex image difference metric. One important consideration when using this module is that of the color space for which the curves are generated. When combined with the spatial filtering modules one can use the opponent color space. If used as an independent module, this type of metric can be applied to the CIE XYZ tristimulus values. Alternatively this type of metric can be used to alter the actual nonlinear compression function of the CIELAB calculations. More discussion on the choice of color spaces will follow in Section 7.6.

Similarly to the spatial localization module, the output of the contrast module can be used to detect whether perceived color differences are a result in overall changes in image contrast. This should be able to determine if differences are a result of changes in white or black points, as well as overall tone-reproduction changes. When used in conjunction with the final color difference equation, this becomes a powerful step towards an ultimate quality model.

7.5 Error Reduction

All of the above modules work as a pre-processing step to an existing “core metric.” For the examples given above this core metric has been the CIELAB color space, and specifically the CIE color difference equations. The general flowchart for this framework is shown below.

Figure 41. General Flowchart for an Image Difference Metric

The input to the metric is two images, and the output is a single image where each pixel represents the magnitude of perceived color difference expressed in terms of the CIE color difference equations such as ΔE_{94} . The output image is often referred to as the error image. This image can provide many insights into the cause and location of the differences between two images, which is often very beneficial when designing imaging systems. Often, though, what is desired is a single number that represents the magnitude of perceived error of an entire image. This can be thought of as the generalized equation shown below:

$$\Delta_{im} = f(image1, image2) \quad (29)$$

where Δ_{im} is the overall image difference, and $f(image1, image2)$ is some form of a color image difference metric. There are inherent dangers in reducing an entire error image into a single number, a term that has been coined “mono-numerosis.” Still, there are indeed some occasions where the calculation of an overall image difference is both necessary and beneficial. This is especially the case when predicting human psychological scales. It is difficult to use an entire error image to predict human behavior. For instance, when examining the perceptibility and acceptability of image compression algorithms humans can either pass or fail a given image. For perceptibility experiments the observers fail an image if they can see a

difference while for acceptability experiments the fail an image if the compression is considered bad. Often times it is desired to map the error image into a single number scale to correlate with the observer perceptibility and acceptability thresholds. The mapping for these two thresholds could potentially be different, as perhaps the maximum error in the difference image correlates to perceptibility, while the average error correlates to acceptability.

As such, there are many different techniques for reducing the information contained in an entire error image into a lower dimensional representation. Perhaps the simplest such methods rely on the “statistics” of the error image itself. These statistics can be the moments of the image, such as the mean and variance of the errors. Higher order moments of the image such as skewness and kurtosis can also be examined. Other statistical approaches are the root-mean squared (RMS) error, as well as the median, maximum and other percentiles.

7.5.1 *Structured Data Reduction*

Other techniques for data reduction are possible as well. While the spatial filtering module is used as an attempt to normalize all spatial frequencies such that the color difference is the same, it is possible that this normalization is not enough. An example of this is shown in Figure 42. The mean color difference between either of the two bottom images and the top image in Figure 42 is identical, when calculated in a standard pixel-by-pixel basis. The image on the bottom left has noise added to it in such a manner that it should be barely perceptible. The image on the right has a large hue shift on one of the bananas. The difference between the green banana image and the original image (top) is very apparent, yet the calculated mean color difference is the same as the additive noise image. This should not be surprising, as the S-CIELAB model was created to correct for just such problems.³³ When the images are run through the image difference framework presented above, the calculated mean image difference for the banana image becomes three times as large as the mean difference of the noise image. While this is very comforting, in that the color image difference metric is performing as it is expected, the mean magnitude of the banana image might still seem small. If we were to look at the full output error map the large banana error (or LBE) would immediately stand out. Computational models that can “look” at the error image for us might provide for much more accurate data-reduction than simple image statistics.

Figure 42. Example of Images with Identical Mean Error. Image on Top is Original, Image on Left has Additive Noise. Image on Right has Green Banana [Image Photographed for MCSL by Eastman Kodak Company].

One possible technique for data reduction is using an error image correlation. An auto-correlation of the error image is a relatively straightforward calculation, as shown below:

$$Auto = \mathcal{F}^{-1} \{ \mathcal{F}\{image\} \mathcal{F}\{image\} \} = \mathcal{F}^{-1} \{ \mathcal{F}\{image\}^2 \} \quad (30)$$

where $\mathcal{F}^{-1}$ and $\mathcal{F}$ are the inverse and forward Fourier transform of the error image. The auto-correlation will serve to further boost errors when they exist in spatially large regions, and suppress errors when they occur in smaller regions. Image statistics can then be used on the auto-correlated error image to obtain a single metric for perceived image difference.

Another possible technique for data reduction could be using image processing clustering techniques. Essentially image clustering can find distinct regions of errors that might be more perceptible, and then weight those clusters more. Many different clustering techniques could be applied to this type of situation.⁶²

7.5.2 *Data Reduction Summary*

The large amount of information contained in an error image often needs to be reduced to a manageable amount of information. This can be accomplished using simple image statistics, or more complicated image-processing techniques such as auto-correlation or image clustering. In the context of the image difference framework, the data-reduction can be thought of as a post-processing module that follows the core metric.

Another potential technique would be to use a hybrid technique that utilizes the strengths of the device-dependent system modeling with the strengths of the modular framework. Often, when modeling an imaging system, system parameters are directly linked with perceptions; linking MTF with sharpness, for example. The various perceptions, or “nesses,” can then be combined to form an ad-hoc image quality model using Minkowski summation. This approach was described in detail in Section 2.1. A similar approach could be used to weight the output of each of the individual modules described, along with the overall color differences, to form a weighted sum of overall image difference. This would allow a type of “importance” weighting to various percepts. The use of various image statistics for experimental prediction are discussed more detail in Section 9.

7.6 *Color Space Selection*

All of the previous modules discussed are either pre-processing or post-processing that occurs on the core metric. For the discussion up until this point the core metric was always assumed to be the CIELAB color space in combination with the CIE color difference equations. This approach has many benefits. CIELAB is an industry and academic standard that is well known, and well understood. Likewise, much research has gone into the formulation of the CIE color difference equations. This research has culminated in the creation of the CIEDE2000 color difference metric.⁶³ It seems very appropriate to piggyback a color image difference metric on top of this historical research.

There is no reason, however, this framework has to rely on CIELAB as its core metric. The modularity of this framework applies just as much to the core metric, as to the other modules. As such, it can easily be replaced with a different metric. There might be good reason to choose a different color space for the core. One important consideration might be the choice of a color appearance space as the core. This could be an important stride towards the creation of an image color space. As the second goal of this research project is to create an image appearance model that is applicable for both image difference and overall appearance predictions. One possible choice for an appearance core could be CIECAM97s⁶⁴ or the newly published CIECAM02.⁶⁵ Using CIECAM02 as a base would have the same benefit as using CIELAB, in that years of historical research has gone into its formulation. There are several drawbacks to the CIE color appearance models, however. They tend to be relatively complicated

models to implement, and would only gain complexity when used in the image difference framework. CIECAM02 also lacks a well-defined color difference equation. While it would be possible to create one using the appearance correlates, the space was not designed for use as a color difference space.

7.6.1 IPT

Another potential candidate for the core metric is the Image Processing Transform (IPT) space published by Ebner.⁶⁶ This space is a relatively simple color space designed for ease of use in image processing applications such as gamut mapping. One of the strengths of the IPT space is its hue-linearity, which is a large improvement over CIELAB.⁶⁶ The general flowchart of IPT is shown in Figure 43.

Figure 43. Flowchart for IPT Color Space

The color space takes input in the form of CIE XYZ tristimulus values. The model assumes these values are adapted to D65. If the image is not displayed under D65 then a chromatic adaptation transform must be used to calculate the corresponding D65 colors. One chromatic adaptation transform that could be

used is the linearized CIECAM97s transform proposed by Fairchild.⁶⁷ The adapted XYZ values are then transformed into cone responses using the following equation.

$$\begin{bmatrix} L \\ M \\ S \end{bmatrix} = \begin{bmatrix} 0.4002 & 0.7075 & -0.0807 \\ -0.2280 & 1.1500 & 0.0612 \\ 0.0000 & 0.0000 & 0.9184 \end{bmatrix} \begin{bmatrix} X_{D65} \\ Y_{D65} \\ Z_{D65} \end{bmatrix} \quad (31)$$

The cone responses are then compressed using a nonlinear function such as that shown below.

$$\begin{aligned} \text{if } L \geq 0 \quad L' &= L^{0.43} \\ \text{if } L < 0 \quad L' &= -(-L)^{0.43} \end{aligned} \quad (32)$$

The function is the same for the M and S responses. Finally, the IPT opponent channels are calculated using a 3x3 matrix transformation, as shown below.

$$\begin{bmatrix} I \\ P \\ T \end{bmatrix} = \begin{bmatrix} 0.4000 & 0.4000 & 0.2000 \\ 4.4550 & -4.8510 & 0.3960 \\ 0.8056 & 0.3572 & -1.1628 \end{bmatrix} \begin{bmatrix} L' \\ M' \\ S' \end{bmatrix} \quad (33)$$

Color appearance correlates can be calculated by transforming the IPT coordinates into a cylindrical space much the same as converting CIELAB into CIELCh. Color differences can be calculated using a Euclidean distance metric on the IPT coordinate.

7.6.2 Color Space Summary

Most of the discussion of the above modules assumes CIELAB to be the core metric from which the color differences are calculated. The modular nature of the image difference framework allows the core metric to be exchanged, if desired. Other possible core spaces could be color appearance spaces such as CIECAM97s and IPT. The simplicity of IPT makes it a very attractive alternative for the core metric. This topic will be revisited in Section 11 with the discussion of an introductory image appearance model.

7.7 Color Image Difference Module Summary

This section has introduced several modules that can be used in the modular image difference framework presented in Section 6. Several pre-processing modules were examined. This includes modules for spatial filtering based on the contrast sensitivity of the human visual system, spatial frequency adaptation, spatial localization filtering, and local and global contrast detection. A module for post-processing was also discussed. This module represents a data-reduction stage for reducing a high-dimensional error image into a reduced representation that accounts for the overall magnitude of error for an image pair. This reduction can take the form of one or more image statistics, such as mean, median, and standard deviation. Finally,

the color-space module at the center of the image difference framework, known as the core metric, was discussed. The concept of changing the core color space will be revisited in a later section.

The modules presented in the above section in no way attempt to represent all possibilities for an image difference metric. There are many other examples that can be used in this type of framework. One prime example would be the orientation tuning and visual masking used by the Color Visual Difference Model.³⁴ It is the hope that this framework will serve as a stepping stone for other researchers to formulate new or refine existing image difference modules.

8 Psychophysical Evaluation

In order to create models of image difference or image quality that accurately predict the perceptions of human observers it is necessary to test the models against experimental data. The following section details a series of psychophysical experiments that have been designed to test various model aspects discussed in the previous sections. There are three main experimental data sets that are used to test the image difference model. These include two softcopy experiments, testing perceived sharpness and perceived contrast. A third hard-copy experiment testing sharpness, graininess, and overall perceived image quality is also described.

8.1 Sharpness Experiment

This first experiment was designed to measure the perception of image sharpness. While only one of the many perceived appearances that make up image quality, it has been noted that sharpness plays a very important role.⁶⁸ Therefore, the study of sharpness presents an ideal starting point towards bridging the gap between spatial and color image difference and quality modeling.

This experiment examines the simultaneous variations of four image parameters: spatial resolution, additive noise, contrast adjustment, and spatial sharpening filters.

8.1.1 Spatial Resolution

Previous research has indicated that for pictorial images, 300 pixels-per-inch at 8 bits-per-pixel is more than adequate for printed color image quality for typical viewing conditions (18 inches, 8x10 prints).⁶⁹ Thus, we focused on three levels of spatial resolution: 300 ppi, 150 ppi, and 100 ppi. Because there were not soft copy displays easily available with such fine spatial addressability we chose to use lower resolution images viewed further away to approximate the same resolution to the visual system. These images were created by sub-sampling a higher resolution image and then using nearest-neighbor interpolation to expand the lower resolution image back to the original size, effectively creating the appearance of larger pixels, for the lower resolution images.

8.1.2 Noise

To examine the influence of additive noise on perceived image quality, four levels of uniform, channel independent RGB noise were created: no noise, 10 digital count, 20 digital count, and 30 digital count noise. Each of the noise levels was uniformly distributed around a mean of 0. A per-pixel noise field was

generated for each level (for instance -10 to +10 for the 20 digital count) and added to the base image. These images were then re-clipped between 0 and 255.

8.1.3 Contrast Enhancement

Three levels of contrast enhancement were used in the experiment. This includes the standard "non-enhanced" level, and two levels of contrast enhancement. The enhancement was performed using sigmoidal exponential shaping functions.

The three levels of contrast (none, exponent 1.1, exponent 1.2) were performed on the independent image RGB values, indicative of a typical image-processing situation.

8.1.4 Sharpening

There exists many image editing tools which allow an end-user the ability to enhance the sharpness of an image, through the use of spatial or frequency filters. One common tool is Adobe Photoshop. In this experiment there were two levels of image sharpening: none, and the Photoshop sharpen filter from version 5.5 on the Mac OS. This filter is a 3x3 Laplacian kernel (2nd derivative with a center of 5) with an additive offset term. This use of a sharpening kernel is meant to simulate post processing one might do on pre-existing images.

8.1.5 Experimental Design

The four different image parameters described above combine to form 72 images, when simultaneous variations are included (3 resolution * 4 noise * 3 contrast * 2 sharpening). The order that the simultaneous variations occur can have a great impact on the resulting images. For this particular experiment a real imaging system, such as a digital camera, was simulated. It is important to stress that this process was not meant to actually model a real imaging system, but was rather a simple approximation of how a real system might behave. The order of processing thus went:

- *Resolution: Similar to resolution of an image capture or output device*
- *Additive Noise: Similar to noise that might occur in image capture*
- *Contrast: Similar to nonlinear processing that occurs in imaging device*
- *Sharpening: Typical user post-processing.*

Figure 44 shows an image matrix representing four image variations, in the order they were applied.

Figure 44. Image Manipulations Performed for Sharpness Experiment

The 72 images were then used in a paired-comparison experiment. In the paired-comparison paradigm, the 72 different images result in 2556 pairs for evaluation ($72 \times 71/2$) for each scene. Four scenes were chosen, golf, cow, man, and bear, and are shown in Figure 45. The golf scene was originally a digital photograph, while the man and bear were PhotoCD scanned silver halide images. The cow image was a computer graphics rendering. The 72 manipulations combined with the 4 scenes resulted in a staggering 10224 image comparisons necessary to get a complete interval scale.

Figure 45. Four Scenes Used in Sharpness Experiment

The pairs of images were displayed on an Apple Cinema digital LCD display, driven by a Power Macintosh G4/450. The 22-inch diagonal display allowed two 4x6 inch images to be displayed simultaneously.

The images were presented on a white background, with a maximum luminance of 154 cd/m². Previous work by Gibson⁷⁰ has shown that LCD monitors are as uniform as high quality CRT displays, and are capable of equally or better color characterization accuracy. To simulate 300-ppi resolution with an 86-ppi device the display was placed at a viewing distance of 5ft, which is approximately 3.5 times a normal print viewing distance of 18 inches. Since the perception being scaled is sharpness it is not suspected that the increase distance has a large effect, though that poses an interesting question for future research. Several of the CSF models take this viewing distance and image size into account, and are discussed in Section 9.1.2. The images presented were 630 by 420 pixels, which subtended roughly 7 degrees of visual angle when viewed at this distance. To facilitate the speed at which pairs could be viewed all 288 different images (72 images x 4 scenes) were loaded into memory. All possible pairs were then randomized and were presented to the observer with random selection between right and left side of the display. The observer was given a left hand and right hand mouse, which they clicked to select their chosen image. With this set up, it was easily possible to present a new image pair in less than .5 seconds. The experimental setup is shown in Figure 43.

Figure 46. Sharpness Experimental Setup

Observers were presented with the rather simple task of choosing which of the two images “appears sharper.” A single session presented 500 pairs of images to an observer. On average, an observer was able to finish a session in 20 minutes. Observers could then choose to continue on for multiple sessions if they desired, or quit after a single session. Since no person could perform all 10224 observations in a single sitting, the experiment was designed to allow an observer to finish a session and resume where they left off at a later date.

8.1.6 Sharpness Results

A total of 51 observers completed over 140,000 observations. Five observers completed all 10224 observations, while the average observer completed roughly one quarter of all the image pairs, approximately 2556 observations.

Thurston's Law of Comparative Judgments, Case V, was used to analyze the results of the paired comparison experiment, and convert the data into an interval z-score scale.⁷¹ Due to vast difference between some of the image pairs, there were several zero-one proportion matrix problems. This was solved using Morrissey's incomplete matrix solution, which uses a linear regression technique to fill in the missing z-scale values.⁷²

The z-score values were then normalized by subtracting the “original” image for each scene. This created an interval scale of sharpness such that any image with a positive score was judged to be sharper than the original, while any image with a negative score was judged to be less sharp. The goodness of fit was tested using the Average Absolute Deviation (AAD), as shown in Equation 34:

$$\overline{p'_{ij} - p_{ij}} = \frac{2}{n(n-1)} \sum_{i>j} |p'_{ij} - p_{ij}| \quad (34)$$

where p'_{ij} is the predicted probability that *Image “i”* is judged sharper than *Image “j”*, p_{ij} is the observed proportion *Image i* is judged sharper, and n is the number of stimuli.

The AAD on the resulting z-scores resulted in an average error of 0.026, or 2.6%. This suggests that the Case V model fits the data well. The complete interval scale of sharpness was averaged across the four scenes, and is shown in Figure 47. The legend shown in Figure 47 shows the ranking of all the image variations, from best to worst. The images are labeled as follows: first, the resolution of the image is listed, followed by the amount of noise, followed by the contrast level, and a sharpness key. For example, image 300+20n+1.2+s is a 300dpi image, with 20-pixel noise, a contrast enhancement of 1.2, and sharpened in Photoshop with a Laplacian edge enhancer. The complete ranking of all the images, along with the corresponding z-scores can be found in Appendix A.

Figure 47. Complete Sharpness Interval Scale, Averaged Across Scenes

These results indicate that 21 images appeared as sharp or sharper than their respective original images, when taken into account the confidence intervals. The error bars shown in Figure 47 indicate the confidence interval resulting from the experiment. This interval can be calculated as follows:

$$CI = \frac{1.38}{\sqrt{N}} \quad (35)$$

Where N represents the number of observations for each image pair. At least 6 images were judged significantly sharper than the original. All of these images had the highest resolution of 300 dpi. This indicates that spatial resolution, or addressability, is of the highest priority for this experiment. The 300-dpi image, with a noise level of 10, a contrast increase of 1.2, and with spatial sharpening was determined to be statistically sharper than all other images. The 300-dpi image, with noise level 20, contrast increase of 1.2, and spatial sharpening was also judged significantly sharper. The effects of each of the manipulations are discussed in more detail below.

The experimental data for all the scenes individually were then examined to see if any scene dependencies were present. For the Cow scene, the Average Probability Deviation calculated was 0.043,

indicating less than 5% error. This indicates that the model used was a good fit for the data. It is important to also note that for the Cow scene, adding noise and increasing contrast to an image was at times able to mask some of the resolution differences between the 300dpi and the 150dpi images. Several enhanced 100dpi images were also judged to appear as sharp as some 150dpi images. The full ranking for the Cow image is given in Appendix A. Another interesting artifact for the cow scene was the effect of spatial sharpening. For most images, the highest ranking images tended to have spatial sharpening, while for the cow this was not the case. Instead, there were many cases where lower resolution images were selected over the spatially sharpened higher resolution image. This suggests that perhaps the edges of the computer rendered cow were already too crisp. This could be because the rendered image had suffered none of the optical degradation that usually occurs in an imaging system, or perhaps it is because the Cow did not appear “natural.” Figure 48 shows the interval scale (y-axis) for the sharpest 30 manipulations in cow scene, normalize to 0 for the original image.

Figure 48. Experimental Results for Cow Scene

For the remaining scenes the Average Probability Deviations were determined to be 0.044, 0.046, and 0.043 for the Bear, Cypress, and Man images respectively. All of these errors were less than 5 percent. This indicates that the Case V model was a good fit for all of the image scenes. For the bear scene in particular, there were several different occasions where a lower resolution image was selected to be sharper than several higher resolution images. This was particularly the case for the 150-dpi vs 300-dpi images. This occurrence was also found in the Cypress images, and less so in the Man images. For all

scenes, the sharpest images had some form of contrast enhancement. Figures 46-48 show the results for the Bear, Cypress, and Man scenes.

Figure 49. Experimental Results for Bear Scene

Figure 50. Experimental Results for Cypress Scene

Figure 51. Experimental Results for Man Scene

To determine whether the combined data analysis masked any particular features evident in the individual scenes, the individual scene Z-scores were plotted against the combined Z-scores. Figure 52 illustrates these plots for two of the scenes, the Cow and Cypress images.

The cow scene fits with the combined data reasonably well with a correlation coefficient of 0.81, though there are some interesting outlying points. All of the data that do not match up well with the combined results involved images that were spatially sharpened. The most noticeable outlying point is the sharpened 300dpi image. While consistently one of the perceptually sharpest images for the other scenes, it was judged not very sharp for the Cow scene.

The other scenes match the combined data rather well, with correlation coefficients of 0.90, 0.96, and 0.96 for the Bear, Man, and Cypress scenes respectively. This analysis indicates that the data for all scenes can be combined for model testing, without worrying about scene dependency. It is important to note that the slope of the lines fitting the data in the above figures is not important, but rather that the data can be fit well with a simple linear equation.

Figure 52. Individual Scene Interval Scale vs. Combined Scene Interval Scale

The individual image variations were then examined to try and gain an understanding of the rules of sharpness perception. All of the z-scores for each particular manipulation were averaged across the combined results as well as individually for each scene. For instance, the z-scores for every image at 300 dpi were averaged to create a scale representing “300 dpi.” This created an average weight, for any given variation. Figure 53 provides a plot of this analysis.

Figure 53. Independent Variation Effects (Average Z-Score)

It is clear from the analysis that spatial resolution, which can be thought of pixel size or addressability, is by far the most important influence on perceived image sharpness. Other interesting "rules" can be interpreted from the results. Enhancing contrast increases the perception of sharpness for all scenes except for the bear. Additive noise increased perceived sharpness, up to a certain amount of Pixel noise, and then decreased sharpness. Spatial filtering had a significant effect of sharpness for all scenes, except the Cow

scene where it decreased perceived sharpness. These effects were most noticeable in the 300 and 150 dpi images. At 100 dpi, the effects were similar, though less distinct.

Clearly the effect of resolution is overwhelmingly dominant. To better understand the effects of the other variations it is necessary to remove the resolution influence. The following series of plots shows the independent variations at each level of resolution. These plots were normalized by averaging all of the z-scores for each attribute at a single resolution level (for instance all images with 20 noise were averaged) and then subtracting the z-score for the average of all the images at that resolution level.

Figure 54. Independent Variation Effects for 300 DPI Images

Figure 55. Independent Variation Effects for 150 dpi Images

Figure 56. Independent Variation Effects for 100 dpi Images

Isolating the other manipulations away from the dominant trait of resolution can reveal several important trends. For the 300 dpi images increasing contrast generally (on average) causes a small increase in perceived sharpness, as illustrated in Figure 54. For the Cow image contrast was very important, as evidenced by the large negative bar for the “0 contrast” manipulation and large positive bars for the increased contrast manipulations. The Bear image displays the opposite relationship with contrast. The 150 dpi images, shown in Figure 55, shows the same general increase in sharpness as contrast increased. The increase in contrast for the 100 dpi images, Figure 56, had almost no effect on perceived sharpness.

Similar relationships can be seen with the other variations. A very interesting relationship is the effect of noise on perceived sharpness. For the 300 dpi images there is a slight increase in sharpness when noise levels of 10 and 20 are added, and slight decrease with noise levels of 30. This is an interesting perception, as it is generally implied that noise and sharpness are inversely related. There could be several reasons for this increase in sharpness perception. The uniform noise has information at all frequencies, including high frequencies. It is possible that the human visual system attributes this additional frequency information to edge information, which is typically associated with high frequencies. Other possibilities include a slight change in local contrast caused by the increased local deviations (noise) while the overall mean of the images remain the same. For the lower resolution images there is a monotonic decrease in perceived sharpness with all levels of additive noise, compared to the images with no noise. This suggests that the frequency of the additive noise, or the appearance of visible noise edges might play a role in the perception of sharpness.

The above figures indicate that the presence of a single strong perceptual influence overwhelms or masks smaller influences. This ties back into the concept of using a Minkowski summation technique

for image quality modeling. By choosing an appropriate summation exponent such a metric could assign much more weight to the most “influential” variable such as resolution. That idea can be thought of as a suppression multivariate model where large error in any variable overwhelms and suppresses the output from any other variable. This suppression model was described in detail Keelan.⁶

8.2 Contrast Experiment

An extensive psychophysical experiment examining the perception and preference of image contrast was performed by Anthony Calabria at the Munsell Color Science Lab.⁷³ This experiment was very similar in design to the Sharpness Experiment, and was based partially on that experiment. Several distinct experiments were performed under the umbrella of this contrast experiment, involving the effects of lightness, chroma, and sharpness manipulations on perceived image contrast. These are discussed in brief below.

8.2.1 Lightness Manipulations

Twenty manipulations of the CIELAB L* channel were performed to test the effect of the lightness channel on perceived contrast. These manipulations include seven sigmoid functions (both “increasing” and “decreasing” contrast), four exponential “gamma” functions, eight linear L* manipulations, and histogram equalization.

8.2.2 Chroma Manipulation

For the chroma experiment, the L* channel from the most preferred image in the lightness experiment was chosen as a base image. This grayscale image had six linear manipulations of CIELAB Chroma (C*) combined with it, resulting in seven images with various amounts of color information, ranging from black-and-white to 120% chroma boosting.

8.2.3 Sharpness Manipulation

Seven unsharp masking functions of various weights were applied to the most preferred L* image from the lightness experiment using Adobe Photoshop. This resulted in eight levels of sharpness, including the original image.

8.2.4 Experimental Conditions

A total of six image scenes were used for the three experiments. These scenes are shown in Figure 57.

Figure 57. Image Scenes Used in Contrast Experiment

Two paired-comparison evaluations were conducted for each experiment, using a setup similar to that used in the Sharpness Experiment. The images were viewed at a distance of approximately 24 inches, and subtended roughly 12 degrees of visual angle, corresponding to 22.5 cycles-per-degree of visual angle (45 pixels per degree). In the first experiment the observers were asked to select their preferred image, while in the second they were asked to select the image with the most contrast. This resulted in interval scales of contrast and preference, as a function of Lightness, Chroma, and Sharpness. The averaged contrast scales (across the 5 pictorial images) are shown in Figures 55-57. The brainscan image was judged differently than the other images, so it was excluded for this study. The remaining 5 images were judged the similar to each other, so they were combined for this analysis. Further details can be found in Calabria⁷³ while the full experimental rankings can be found in Appendix A.

Figure 58. Contrast Scale Resulting From Changes in Lightness

The legend in Figure 55 shows the rank of the various images, from “least contrast” to “most contrast.” The image titled “gma_0.900” is the image raised to a power (gamma) of 0.90. If the image title begins with “inc” or “dec” that implies an increase or decrease in with sigmoid function, and if the image title begins with “lin” it is a linear shift in the black level. Applying an increasing sigmoid function does increase perceived contrast, as does applying a power function greater than 1.0. Figure 59 shows the general trend for increasing and decreasing sigmoid and gamma lightness functions. One interesting note is that the increasing sigmoid always appears more to be of higher contrast than the original (labeled gma_1.00 in the legend of Figure 58), The images with less of an increase in sigmoid were determined to be of higher contrast than those with a greater increase. In addition to the nonlinear manipulations, moving the black level linearly to a lower value increases the perception of contrast, while raising the blackpoint decreases it.

Figure 59. Iconic Representation of Sigmoid and Gamma Lightness Functions

Figure 60. Contrast Scales Resulting From Changes in Sharpness. Sharpness Level Represents Percent Unsharp Mask Addition

The effect of sharpening on perceived contrast varied linearly with sharpness level. Essentially higher “sharpening” resulted in an increase in perceived contrast. This correlates very well with the sharpness experiment described in Section 8.1. In that experiment, increasing contrast resulted in an increase in perceived sharpness. The results from this experiment and the sharpness experiment suggest there is a high correlation in the perceptual relationship between contrast and sharpness.

Figure 61. Contrast Scales Resulting From Changes in Chroma

The relationship between chroma and perceived contrast, as illustrated in Figure 61, is generally linear. As chroma increases so too does contrast. The one exception to this rule is the image with a contrast scale of 0.2. This image looks almost achromatic, but is actually judged to be significantly of less contrast than the black-and-white image. From these results it can be said that contrast is dependent on both achromatic and chromatic information.

Complete analysis of these experiments can be found in detail in Calabria,⁷³ with summary z-scores shown in Appendix A.

8.3 Print Experiment

A joint hard-copy experiment was performed by researchers at RIT and at Fuji Photo Film in Japan. This experiment was first designed and implemented at Fuji, and then subsequently repeated at RIT. The experiment consisted of two scenes, with a series of four manipulations on each image. The two scenes, Portrait and Ship, are shown in Figure 62.

Figure 62. Image Scenes Used in Print Experiment

The print experiment was designed to simulate several aspects of digital CCD camera design, and as such there were a series of manipulations that correspond to what might happen in an actual camera. One of the manipulations is simulated ISO speed, corresponding to the physical pixel size, approximating “film grain.” This is illustrated with a close-up of the Portrait image for ISO speeds of 320 and 1600, as shown in Figure 63.

Figure 63. Simulated ISO Speed Corresponding to 320 (left) and 1600 (right)

The next manipulation was a frequency cut-off filter, often used to prevent aliasing from a regularly gridded CCD camera. Two frequency cut-off filters were tested, a rectangular shaped filter, and a diamond shaped filter, which was a rotated version of the rectangular designed to maintain frequency information in the horizontal and diagonal direction. An iconic example of the 2-D filter is shown in Figure 64. The frequency cut-off is at half Nyquist, and is meant to simulate a typical anti-aliasing type filter.

Figure 64. Iconic Representation of Frequency Cut-off Filter

Two levels of frequency boosting, or sharpening, were then applied. These filters can be thought of as similar in nature to the spatial localization filter described in 7.3.3 and shown in Figure 37. Finally two levels of additive noise were added to the images. The two each of speed, frequency cut-off, sharpening, and additive noise yielded a total of 18 manipulations.

8.3.1 Print Experimental Setup

The digital images were printed on a Fuji photographic printer at a size of 8x10 inches, for use in a rank-order experiment. The observers were asked to rank each of the images along three dimensions: sharpness, graininess, and overall print quality. The prints were viewed in ISO conditions of D50 simulators at approximately 1000 cd/m². The observers were allowed to handle the images, and were told to use a “normal” viewing distance of 12 inches. At this condition the images subtended approximately 30 degrees of visual angle, or 35 cycles-per-degree of visual angle (70 pixels-per-degree). A total of 20 observers participated in the RIT experiment; 13 males and 7 females. There were 13 observers considered experienced in this type of judgment, while 7 were considered naïve. A total of 25 observers participated at Fuji, all male with experience in this type of judgments. The rank-ordered data were converted into z-scores using Thurstone’s Law of Comparative Judgments.⁷⁴ These z-scores represent interval scales for sharpness, graininess, and quality. The complete results of both the Fuji and RIT experiments can be found in Appendix A. The RIT results are shown in Figures 65-70.

Figure 65. Interval Scale of Sharpness for Ship Image

Figure 66. Interval Scale of Graininess for Ship Image

Figure 67. Interval Scale of Quality for Ship Image

Figure 68. Interval Scale of Sharpness for Portrait

Figure 69. Interval Scale of Graininess for Portrait

Figure 70. Interval Scale of Quality for Portrait

The actual rank order of the image manipulations are shown in the legends of Figures 61-66. The image name begins with the ISO speed, and is followed by the frequency boost, cut-off filter design, and noise addition (e.g. iso1320_freq1_diam_noise2). The rank-order data is scaled from highest to lowest for the scaled attributes. What stands out most from these data is the overwhelming influence of ISO speed on all the scales. The ISO 300 images were universally judged to be higher in quality and sharpness, and lower in graininess than the ISO 1600 images. Since the z-scores are arbitrary interval scales, they can be normalized by the addition or subtraction of any integer. This is illustrated in Figure 71, which shows the overwhelming influence of ISO speed on perceived quality for the Portrait image. In this situation the z-scores were normalized such that the lowest quality image have a z-score of 0, and increasing z-scores correspond to an increase in quality. The identical image manipulations for each ISO speed are plotted side by side in Figure 71, and are arbitrary labeled 1-9 for convenience. While a general trend can be seen for all the manipulations, the trend is not statistically significant and is overwhelmed by the ISO speed.

The experiment precision was determined by directly comparing the resulting interval scales between the RIT experiment and the Fuji experiment. This comparison is shown in Figure 72 for the Ship image, and Figure 73 for the Portrait.

Figure 71. Normalized Z-Scores Illustrating the Significance of ISO Speed.

Figure 72. Comparison Between RIT and Fuji Data for Ship Image

Figure 73. Comparison Between RIT and Fuji Data for Portrait Image

In general the RIT data and the Fuji data match up very well, showing a high correlation coefficient. The one notable exception is the scaling of sharpness for the Portrait image, illustrated by the upper right image in Figure 73. For this particular attribute the RIT and the Fuji data do not match well at all, and the correlation appears almost random. Since the remaining scales do match up well this could indicate the difficulty in judging sharpness for the portrait image, or perhaps the different judgment criteria from the two populations. The Fuji group was made up entirely of male expert observers, while the RIT group was a mixture of male and female, naïve and experts. These two populations might have differing internal standards of sharpness for portraiture images.

It is also important to examine the scene dependency from this experiment. It should be noted that in general both the Sharpness and Contrast experiments produced scales that were relatively scene independent. The noticeable exception was the Brainscan image in the Contrast experiment. To determine the scene dependence the z-scores for each particular manipulation were plotted against each other for the ship and portrait scene. The z-scores for the three different scales were examined, and are shown in Figure 74.

Figure 74. Scene Dependence for Print Experiment

For the quality and graininess experiments there seems to be little scene dependence. The same cannot be said about the sharpness experiment, which shows considerable difference between the portrait image and the ship image. Again, this seems to indicate an inherent difficulty for observers to judge sharpness, or a difference in the sharpness criteria across populations. The scene dependency plots, particularly for the image quality experiment (upper left in Figure 74) show two distinct clusters of data corresponding to the ISO speed rating of the images. This strongly suggests an overwhelming influence of ISO speed on the perception of overall image quality.

Figure 75. Z-Score Values of Image Quality Experiment

Figure 75 shows the image quality z-score values for the ship image, represented by the y-axis. The x-axis represents a simple nominal scale representing different image manipulations. The specific manipulation labels are not important as these plots are meant to illustrate the importance of ISO speed. There are two distinct groups shown, corresponding to the ISO 300 speed images, and the ISO 1600 images. This indicates that the ISO speed is by far the most important attribute used to judge overall image quality. Similar plots are shown in Figure 76 for the sharpness and graininess scales.

Figure 76. Z-Score Values for Sharpness and Graininess Experiments

This same trend can be seen in the graininess scale, with the high-speed ISO 1600 image being considerably grainier than the ISO 300 images, for the ship image. The sharpness scale does follow this same general trend, though several ISO 1600 images were judged to be as sharp as some of ISO 300 images. Examining the quality scale for the Portrait image shows the same large distinction between the two ISO speeds, as shown in Figure 77.

Figure 77. Z-Score Values for Image Quality Experiment (Portrait)

The graininess and sharpness scales for the portrait image reveal this same story. For both of these scales the ISO speed was judged to be the most important indicator, as shown in Figure 78.

Figure 78. Z-Score Values for Sharpness and Graininess Experiments (Portrait)

This trend is very interesting, considering general difficulty observers had with judging sharpness for the portrait image and the discrepancy between the RIT and Fuji groups. The overwhelming influence on ISO speed for the print experiment is similar in nature to the influence of resolution in the Sharpness Experiment, as discussed in Section 8.1.6. This again indicates the possible ability for a single image attribute to mask, or overwhelm, other less important attributes.

8.4 Psychophysical Experiment Summary

Three psychophysical attributes were discussed in this section. These experiments scaled several image attributes such as sharpness, contrast, graininess, and also overall image quality. These experiments are all interesting in and of themselves. The data collected yields itself to almost unlimited analysis. In the following section these data will be used to test the color image difference framework, as well as several of the individual modules previously discussed.

9 Image Difference Framework Predictions

This section discusses the use of the color image difference framework to predict the results of the psychophysical experiments described in Section 8. While these experiments were not designed to directly scale perceived image difference, they were designed to scale individual image attributes. In order to scale any given attribute for a pair of images, an observer must first be able to see a difference between those images. Larger attribute differences should correlate with a larger perceived image difference.

The image difference framework is designed to predict the magnitude of perceived color difference between an “original” and a “reproduction.” The three experiments described above create interval scales of sharpness, contrast, graininess, and quality based upon either rank-order or paired-comparison analysis. The data can be analyzed using an image difference metric by computing the color difference between every image and a single “original” image. This difference scale can be compared to the experimental data by calculating the interval scale difference between every image and the same original. This results in both positive and negative z-scores, where positive implies the scaled attribute is “greater” than the original, while negative values imply “less” of the given attribute. For the sharpness experiment this means that all images with a z-scale greater than 0 were judged to be sharper than the original.

It is important to note that an image difference metric is incapable of predicting whether an image is deemed to be “more-sharp” or “less-sharp,” only that there is a perceived difference between the images. This means it is essentially capable of predicting the magnitude of difference but not the direction, or cause, of the difference.

9.1 Sharpness Experiment

The experimental sharpness interval scale can be used to evaluate the performance of the various modules in the image difference framework. The sharpness scale can also be thought of as a difference scale, whereas an increase or decrease in perceived sharpness between two images should be directly related to the perceived difference between the two images.

The input to the color image difference framework is two images. The Sharpness experiment involved an “original” image, with 71 different manipulations applied to this image. Thus the image difference metric will calculate the perceived difference between this original image and all 71 manipulations.

9.1.1 Baseline

To begin the image difference framework analysis it is important to start with a baseline calculation. The baseline should be the “core metric” by itself, following the framework presented in Section 7.6. For the following examples that core metric is taken to be CIELAB along with the CIE ΔE_{94} color difference equations. The CIEDE2000 color metric provided similar results for this application. A pixel-by-pixel color difference calculation for the original image and the 71 manipulations was performed. The calculation involves converting the images into CIE XYZ tristimulus coordinates, and then converting those coordinates to CIELAB values based. These calculations were performed based upon the characterization of the display monitor. Once the images are in CIELAB color space, a color difference calculation is performed for each pixel location between the original and reproduced image. The result of the pixel-by-pixel calculation is an error image, where each pixel is considered an independent stimulus. An example of two pixel-by-pixel error images is shown in Figure 79.

Figure 79. Example of Two Pixel-by-Pixel CIE ΔE_{94} Error Images

To compare the error images against the experimental sharpness scale the error images must be reduced in dimension. The post-processing data reduction for the following example is the mean color difference across the entire error image. Thus the core metric prediction of the experimental sharpness scale is shown in Figure 80. This serves as the starting point for all model evaluation. If the modules do not improve upon the general performance of the core, then there is no need to add the complexity.

Figure 80. Core Metric Error Prediction vs. Experimental Sharpness Scale

There are several important features illustrated in Figure 80 that need to be explained in greater detail, as the same qualitative analysis will be used for all of the modules in the image difference framework. A more formal quantitative analysis is performed in Section 9.1.9. The sharpness scale was normalized such that the original image has a perceived sharpness of “0” units. This was accomplished by taking the z-score difference between every image and the original image. Thus any image that has a positive scale value is perceived to be sharper than the original, while any image that has a negative scale value is perceived to be less sharp. The general form of the image difference model is incapable of determining whether a difference will result in an increase or decrease in sharpness, so it cannot predict the direction, only the magnitude. The means that the ideal plot in Figure 80 would be a “V” shaped plot with the point at the origin [0,0], as illustrated by the lines drawn on the figure. An ideal prediction would also have the same slope for both sides of the “V,” indicating that both positive and negative changes in an attribute are induced by the same perceptual color difference.

Clearly it can be seen that the core metric does not produce anything resembling a “V” shape. This indicates that the core metric itself, otherwise known as a pixel-by-pixel color difference, does not adequately predict the experimental data. This should not come as a surprise since the color difference equations were designed to predict differences of simple color patches and not spatially complex image stimuli.

9.1.2 Spatial Filtering

That the CIELAB color difference equations do not work well for complex, spatially varying stimuli should not be a surprise. The S-CIELAB spatial model was created for just such reasons.³³ This section will examine the effect of the spatial filtering module on image difference prediction. Figure 82 shows S-CIELAB model predictions, using a core-metric of the CIE ΔE_{94} color difference equations. This calculation follows the path of Figure 16; the images are transformed to XYZ and opponent spaces using display characterizations, the opponent channels are convolved with the three separable kernels, and then converted back to XYZ values. These XYZ values are converted to CIELAB coordinates, and a pixel-by-pixel color calculation is performed to get an output error image. An example of a S-CIELAB calculated error image and the difference between that S-CIELAB and a standard pixel-by-pixel calculation is shown in Figure 81. The effects of blurring the high frequency information is illustrated when examining the calculations of S-CIELAB and pixel-by-pixel calculations, as shown by the right image of Figure 81.

Figure 81. Example of S-CIELAB Error Image. Image on Left is S-CIELAB Error Image. Image on Right is Difference Between S-CIELAB and Pixel-by-Pixel Calculations.

Figure 82. S-CIELAB Model Predictions vs. Experimental Results

Again the mean of the S-CIELAB error image is plotted against the experimental sharpness scale. Somewhat surprisingly, the S-CIELAB model does not improve the prediction over than the standard color difference equations, as evidenced by the random scattering of points in Figure 82. This suggests that the S-CIELAB convolution spatial filters might not be adequate for all purposes. This indicates the need to test other more flexible spatial filters. Error images calculated by replacing the S-CIELAB filters with the Movshon three-parameter contrast sensitivity functions, as discussed in Section 7.1, are shown in Figure 83. For these calculations the DC component of the filters were clipped to 1.0, essentially turning the filters into low-pass filters. The image on the right of Figure 83 shows the difference between this calculation and the pixel-by-pixel calculation of the core-metric. Notice most of the differences are in the high-frequency information. Figure 84 shows the experimental predictions using these spatial filters.

Figure 83. Error Image Calculated with Low-pass Movshon Filter (left) and Difference in Calculations from Base Pixel-by-Pixel (right)

Figure 84. Model Predictions Using Movshon CSF

The CSF filters can also be normalized to 1.0 at the DC component, resulting in filters that both modulate and enhance specific spatial frequencies. This serves to actually enhance errors where the human visual system is most sensitive, as well as modulate the high spatial frequencies. An example error image using these filters, along with the difference between this error map and the pixel-by-pixel calculations is shown in Figure 85. The mean color differences found using a frequency-enhancing filter are shown in Figure 86.

Figure 85. Example Error Image Calculated Using Frequency Enhancing Filter (left) and Difference Between Error Image and Pixel-by-Pixel Calculation (right)

Figure 86. Model Predictions Using Movshon CSF with Frequency Enhancement

The CSF filtering that is closer to actual human response, along with the boosting of image difference information around four cycles-per-degree of visual angle show a general improvement over standard S-CIELAB. Figure 86 hints at the desirable “V” shaped trend. This trend will be examined more quantitatively in Section 9.1.9. The boosting of lower frequency information can be seen in the right-hand

image of Figure 85, which shows differences in the error calculations for both the higher frequencies as well as the lower frequencies.

9.1.2.1 Complex Contrast Sensitivity Functions

It is important to understand whether it is possible to gain a further improvement by using one of the more complicated CSF functions described in Section 7.1. These more complicated models include the Barten and Daly CSFs. The error images, along with the differences between those images and the pixel-by-pixel calculations are shown in Figure 87. The error images for these two CSFs look similar, as do the differences in the error images and the pixel-by-pixel calculations. The Barten CSF shows more boosting of lower frequency information, as seen in the lower-right image of Figure 87.

Figure 87. Error Images Calculated with Daly CSF (top left) and Barten CSF (bottom left) along with Differences Between Error Image and Pixel-by-Pixel Calculation (right)

Model predictions of the sharpness experiment using the Daly CSF and the Barten CSF, along with CIE ΔE_{94} , are shown side-by-side in Figure 88. These predictions show similar trends to each other, which would be expected from the similarity between the error images shown in Figure 88.

Figure 88. Model Predictions Using Daly and Barten CSF with Frequency Enhancement

The model predictions using the Movshon CSF and the Daly CSF are virtually identical in shape. The predictions using the Barten CSF are also very similar. This extent of this similarity can be verified by plotting the model predictions against each other, as shown in Figure 89.

Figure 89. Model Predictions Using Various CSFs

The near linear relationship between the three contrast sensitivity functions indicates that the more complicated CSFs did not improve predictions for this particular experiment. The significantly simpler three-parameter model is adequate for these viewing conditions.

9.1.3 Spatial Frequency Adaptation

Recall that spatial frequency adaptation serves to shift and boost the general shape of the contrast sensitivity function. This was illustrated above in Section 7.2.1. The two spatial frequency adaptation techniques discussed can be evaluated using this experimental data. An example error image for the natural scene adaptation based on the $1/f$ assumption calculated using the Daly CSF, along with the

difference in prediction from the original Daly error image and the pixel-by-pixel calculation is shown in Figure 90.

Figure 90. Error Image Calculated Using Natural Scene Spatial Frequency Adaptation (top) and Differences Between Pixel-by-Pixel Calculations (bottom left) and CSF with no Adaptation (bottom right).

Figure 91. Natural Scene Adaptation Module Using Daly CSF

The Daly model is presented here because it gave similar performance to the other CSF models, but adds an anisotropic nature that the other models do not have. The anisotropic shape will be more sensitive to the image-dependent frequency adaptation if the images themselves are not isotropic in their frequency content. More details can be found in Johnson.⁷⁵ The error images generated with the natural scene adaptation appear similar to the error images calculated using just the CSF, as evidenced by the top image in Figure 90. The actual differences between the predictions, shown in the bottom right of Figure 90, illustrates that most of the predictions are actually different. The effect of these differences on the experimental predictions is shown in Figure 91. The general “V” shape trend is improved upon, resulting in a tighter distribution. Again, more quantitative analysis is presented in Section 9.1.9. This indicates that spatial frequency adaptation can be a valuable module, even with a simple natural scene assumption.

The more complicated image dependent spatial frequency adaptation model can be used in a similar manner. The calculated error image, along with the difference between the pixel-by-pixel calculation and the standard CSF calculation is shown in Figure 92. The differences between the image dependent error image and the standard CSF are more localized towards high frequencies, as shown in the bottom right of Figure 92.

Figure 92. Error Image Calculated Using Image Dependent Spatial Frequency Adaptation (top) and Differences Between Pixel-by-Pixel Calculations (bottom left) and CSF with no Adaptation (bottom right).

The results of the model prediction, again using the Daly CSF along with the image dependent adaptation are shown in Figure 93.

Figure 93. Model Predictions Using Image Dependent Frequency Adaptation

The image dependent spatial frequency adaptation model shows a tight grouping when used in combination with the Daly CSF. This suggests that the frequency content of the images themselves does not necessarily follow the $1/f$ assumption. Further quantitative examination of the model predictions follows in Section 9.1.9. The image dependent adaptation also separates the experimental sharpness scale into three distinct groups. These groups correspond to resolution, which makes sense since the image dependent adaptation would be able to “pull out” the resolution information contained in the image itself. More details regarding this type of analysis are described in Section 10.

9.1.4 Spatial Localization

The module for spatial localization serves to model the human visual systems ability to detect edge information, as described in Section 7.3. This module is tested against the experimental data using the simple Sobel method described above. This module is used in conjunction with CSF filtering. The general process is to convert the images to CIE XYZ tristimulus values, and then opponent space representations using the display device characterization. The opponent images are then filtered using the CSF. The spatial localization module is applied to the filtered opponent images, which are then transformed back to XYZ and then CIELAB coordinates. A pixel-by-pixel color difference is then taken between the original and reproduced images. For the viewing conditions of the experiment, the Sobel kernel corresponds to enhancing a region centered on 30 cycles-per-degree of visual angle. An example error image calculated using this process, along with the difference in error image calculation in a pixel-by-pixel procedure is

shown in Figure 94. The difference between the pixel-by-pixel calculation and the model predictions with the CSF and Sobel kernel is quite large, as shown on the right-hand image in Figure 94. These differences includes both low frequencies from the enhanced CSF, and high frequencies from the Sobel kernel.

Figure 94. Error Image Calculated Using Sobel Spatial Localization (left) and Difference From Pixel-by-Pixel Error Calculation (right)

Figure 95. Spatial Localization Model Prediction With Movshon CSF

The experimental predictions of this edge detection module, cascaded with the Movshon CSF are shown in Figure 95. The spatial localization module shows an improved “V” trend when predicting the experimental results. More quantitative analysis of this improvement is provided in Section 9.1.9. The perception of sharpness is often considered to be correlated with the existence of high frequency edge information, so the addition of a module designed to predict the differences of high-frequency information should improve the prediction of sharpness. Similar results are obtained using a cascaded CSF approach with a Gaussian filter tuned to 20 cycles per degree, with width of 10 cycles-per-degree, as

described in Section 7.3.3. An example error image using this approach, along with the differences from pixel-by-pixel and the Sobel calculations is shown in Figure 96.

Figure 96. Error Image Calculated Using Cascaded CSF and Spatial Localization (top) and Differences Between Pixel-by-Pixel Calculation (bottom left) and Sobel Edge Detection (bottom right)

The error image calculated with the cascaded CSF and Gaussian frequency boost shows very specific edge enhancement, as illustrated by the bottom left image in Figure 96. This edge detection is localized to a more specific range of frequencies than the Sobel kernel, as shown by the difference between the Gaussian predictions and the Sobel predictions in the bottom right image of Figure 96. The experimental predictions of the cascaded localization module are shown in Figure 97. While the predictions are not as closely grouped as the Sobel kernel, this type of filter is much more flexible with regards to viewing conditions. The Gaussian filter also appears to better predict the “V” shape of the images judged to be more sharp (positive sharpness scale values.) Further quantitative analysis of this prediction is presented in Section 9.1.9. Identical results to the Sobel module can be obtained by cascading the Fourier transform of the Sobel kernel with the CSF functions.

Figure 97. Spatial Localization Using Gaussian Edge Enhancing

9.1.5 Local and Global Contrast Module

The local and global contrast module is designed to predict differences caused by changes in contrast between two images. The module is applied in conjunction with the CSF module by first transforming the images into CIE XYZ and opponent color space representations through the display device characterization. The opponent images are then filtered using the CSF filters. The achromatic channel of the filtered opponent images is then passed to the contrast module. A family of tone reproduction curves are generated from a low-passed version of the achromatic channel, and these curves are applied to the achromatic channel. The images are then transformed back to XYZ and CIELAB coordinates and a pixel-by-pixel error image is calculated between the original and the reproduction. An example error image, along with its difference from the pixel-by-pixel error calculation is shown in Figure 98.

Figure 98. Error Image Calculated Using Contrast Module and CSF Filtering (left) and Differences Between Pixel-by-Pixel Calculation (right)

The contrast module is designed to predict changes in both local and global contrast. Local-contrast changes can be represented by changes in edge information, as illustrated by the right-hand image in

Figure 98. The global contrast changes are characterized by errors across the entire image. The experimental predictions for the local and global contrast module cascaded with the Movshon CSF are shown in Figure 99.

Figure 99. Local Contrast Module Prediction Using Movshon CSF

This module does not show evidence of the tight distribution of points shown illustrated by the spatial localization module. It does however show an increase in the calculated error for the “positive” images, or those images deemed sharper than the original image. The strength of the local contrast module becomes evident when all of the modules are cascaded together.

9.1.6 Cascaded Model Predictions

All the model predictions until this point have been analyzing the individual modules independently, or in conjunction with the CSF filters. This helps in the development and evaluation of each individual module. To maintain the flexible nature of the framework it is important that the individual modules do not interfere or suppress each other, essentially creating predictions that are worse when used in conjunction with each other than when used independently. An example error image is shown in Figure 100 for calculations performed with the Daly CSF, Image-dependent spatial-frequency adaptation, Gaussian spatial localization, and the local-and-global contrast module. This error image is shown along with the differences in prediction using a pixel-by-pixel calculation. The difference between the pixel-by-pixel calculations and the cascaded model calculations show up for almost every pixel in the image, as shown by the right-hand image in Figure 100. These differences are still most prevalent in the high-frequency edges, where the CSF modulates information based upon properties of the human visual system. Figure 101 shows the experimental prediction of the image difference metric when all of the modules are used together.

Figure 100. Example Error Image From Cascade Image Difference Modules (left) and Differences between Pixel-by-Pixel Calculations (right)

Figure 101. Cascaded Image Difference Modules

The individual modules do not appear to directly interfere with each other, as there is no immediate evidence of decreasing performance. Empirical metrics to test the goodness of fit of each of the individual modules, as well as the cascade model will be discussed further in Section 9.1.9.

9.1.7 Color Difference Equations

All of the plots presented thus far in Section 9.1 have shown the mean of the color difference error image, where the color difference equations selected are the CIE ΔE_{94} . This choice of color space, and corresponding color difference equations are eminently flexible. The use of a different color space, the IPT space, will be discussed in further detail in Section 11.1. The calculation of CIE ΔE_{94} is relatively straightforward compared to the traditional CIE ΔE_{ab}^* equations, while offering significant improvements in color difference predictions.⁷⁶ It is interesting to determine if the added complexity of CIEDE2000 provides a general improvement to model prediction. Figure 102 shows the normalized differences in

prediction of the sharpness data-set using just the simple Movshon CSF with no other modules, between CIE ΔE_{94} and CIEDE2000. These differences have been normalized so that zero difference is represented by gray, while an increase in error prediction for CIEDE2000 results in black pixels and a decrease by white pixels. The similarities between the two color difference equations for most colors is illustrated by the mostly gray color of Figure 102.

Figure 102. Normalized Differences Between CIE ΔE_{94} and CIEDE2000 Predictions

Figure 103. Model Predictions Using CIE ΔE_{94} and CIEDE2000

The experimental predictions of the two color difference formulae are plotted in Figure 103. The two plots are very similar indicating that the two color difference formulae behave similarly in the application of an image difference metric. It should be stressed that this section is not attempting to show which color difference formula can better predict the sharpness experiment, but rather attempting to determine the relationship between the two formulae. This relationship can be examined further by plotting the predictions against each other, as shown in Figure 104. The two color difference formulae predictions are highly correlated, as evidenced by the correlation coefficient of 0.99. They begin to differ slightly at higher color differences, though that difference is rather minimal. One interesting note is that CIEDE2000

color differences are of slightly lower magnitude than their corresponding CIE ΔE_{94} calculations, as evidenced by the slope of 0.77 of the trendline. From this analysis it seems that either of the color difference calculations can be used with similar results. The prediction of the more complex CIEDE2000 equations might improve for images entirely spanning the portion of color space where CIEDE2000 improves upon CIE ΔE_{94} , but for most images these improvements appear to average out.

Figure 104. CIEDE2000 Model Predictions vs. CIE ΔE_{94}

9.1.8 Error Image Reduction

As mentioned above, all of the model predictions mentioned so far have been calculated by taking the overall mean of the color difference error image. The error image itself is a high-dimensional representation of overall perceived error, but by itself it cannot be used to predict the experimental results. This high-dimensionality needs to be reduced in order to enable model testing. While taking the mean is the most straightforward image statistic to use, it is of interest to determine if any other simple statistics are better correlated with the experimental sharpness experiment. Other statistics besides the mean can include median, standard deviation, maximum, and other higher moments such as skewness and kurtosis. While mean error might correspond to overall magnitude differences, it is thought that other statistics might better represent how an observer perceives image differences. For instance, when determining perceptibility between two different images an observer might look for a single area that is different. The maximum error between two images might better correlate for this type of detection. For acceptability experiments, however, observers might examine the overall or average difference between two images to determine if those differences are acceptable. In these cases the mean or median might be more appropriate. Since we do not know what internal criteria the observers used when scaling sharpness it is important to examine many of the possible image statistics.

Figure 105. Mean and Median Color Difference Predictions

Figure 105 shows the mean and median color differences plotted side-by-side. Qualitatively the mean of the image difference produces more of the desirable “V” shape as evidenced by the tighter grouping.

Other higher-order percentiles, sometimes referred to as quantiles, show similar behavior to the median.

Figure 106 shows the standard deviation and the maximum color difference plotted side-by-side.

Figure 106. Standard Deviation and Maximum Color Difference Predictions

The standard deviation shows similar predictions to the mean, while the maximum illustrates some interesting properties. Neither of these statistics correlates as well with the experimental sharpness scale as the mean does, but we begin to see some differentiation of groups in these plots. This indicates that these statistics might not be ideal for predicting overall image difference, but might be useful for predicting the cause of these color differences. This subject is revisited in Section 10.

9.1.9 Metrics for Model Prediction

The results described in the previous are qualitative in nature and are meant to describe if the image difference modules are capable of capturing the general trend of the sharpness experiment. This type of analysis is important as it allows us to determine if an overall image difference perception is related to the

perception of sharpness. Modeling of this nature has not been performed before, and so it is first necessary to gain a general understanding of the relationships between image differences and sharpness perception. That the model behaved similar to what was expected indicates that overall perceived difference can be related to a complex perception such as sharpness. Due to the very complex multi-variate nature of the sharpness experiment, the image difference metric is not capable of fully predicting the results, as indicated by the general spread of the model predictions. There is also inherent variation from the experimental procedure. Therefore the best a model can hope for is to predict the experimental results within this variation. It is important to be able to predict the general experimental trend, illustrated by the “V” shape of the above plots. However, it is also often desirable to have an empirical test of the model illustrating the relative strength of the predictions. This “V” shaped nature indicates that the model is predicting two general trends; the images judged to be sharper and the images judged to be less sharp. Thus a linear regression on these two groups should indicate how well the model predictions correlate with the experimental sharpness scale. The slope of the regression line is generally unimportant in this type of analysis, as the z-scores are arbitrary interval scales. What is of interest are the correlation coefficients as well as the intercept. The intercept is important, as in an ideal situation it would converge to zero, indicating a pair of images that are imperceptibly different. If the model was able to converge upon zero the model would be perfectly able to predict difference thresholds. While this might be an ideal goal, in reality this might be an unrealistic. Traditional color difference formulae are able to predict differences between physical stimuli that are imperceptible to the human observer. As such it is necessary to assign “threshold” ellipses to color difference formulae. These ellipses represent the range of color differences that are below threshold. Also, do to the inherent nature of the Euclidean distance metric in color difference calculations it is not possible to have any errors below zero. This suggests that the model is not normally distributed around zero. This lack of normality near zero makes it difficult to have a predicted zero intercept. This relationship will be discussed in further detail below.

It is important to first establish the empirical fits of the baseline metrics. This enables us to determine whether the need for additional modules is necessary, or if the additional modules do not improve prediction. This section outlines the empirical analysis of the individual modules. The plots of the S-CIELAB predictions as well as the simple modified CSF predictions are shown in Figure 107. These two predictions can serve as the baseline for state-of-the-art image difference models (capable of predicting magnitude differences).

Figure 107. Strength of Prediction, S-CIELAB & Movshon CSF

From these plots we can see that the standard S-CIELAB model has a correlation coefficient (represented by the R^2 calculation) of 0.02 and an intercept of 3.03 for the less sharp images, and a correlation coefficient of 0.32 and intercept of 3.00 for the sharper images. These correlations are very low, and almost random in nature. By changing filters from the defined S-CIELAB kernels to the Movshon CSF the correlations improve to 0.41 and 0.47 respectively, while the intercepts are 1.51 and 1.55. While still not high, these correlations are an improvement over the tradition S-CIELAB. These correlations provide a baseline for each of the modules performance gains or detriments. Clearly the S-CIELAB model is barely correlated with the data, while the Movshon CSF shows a increase in prediction performance. The two more complicated CSFs, from Barten and Daly can also be examined in this manner.

Figure 108. Strength of Prediction, Barten & Daly CSF

These models have similar correlation coefficients to the Movshon model, with the Barten CSF showing slightly poorer performance. The correlation coefficients are summarized in Table 4. This type of analysis shows that simple CSF filtering, while an improvement, is not fully capable of predicting experimental results. Without further refinement it is probably not possible to use this type of model as a replacement for actually psychophysics.

This analysis can be applied to the remaining modules to determine the relative improvements or degradation of their predictions. The image difference predictions for each of the individual modules and their empirical fits are shown in Figure 109.

Figure 109. Independent and Cascaded Model Predictions

From these plots it can be seen that each of the modules does increase the correlation of the predictions, with some illustrating greater improvement than others. Cascading all of the modules together into the

“complete” image difference metric shows the best prediction, indicating that the sum of the parts is in fact greater than the individuals. Table 4 shows the correlation coefficients and intercepts for all of the independent modules, as well as for the cascaded model.

Table 4. Goodness of Fit for Model Predictions

	<i>Negative Sharpness</i>		<i>Positive Sharpness</i>	
Module	R ²	Intercept	R ²	Intercept
S-CIELAB	0.02	3.03	0.32	3.00
Movshon CSF	0.41	1.51	0.47	1.55
Barten CSF	0.31	1.44	0.41	1.52
Daly CSF	0.41	1.47	0.43	1.52
1/f Spatial Adaptation	0.64	1.68	0.53	1.60
Image Dependent Adaptation	0.82	2.77	0.39	1.83
Spatial Localization	0.80	2.28	0.29	2.31
Local Contrast	0.38	1.33	0.47	1.33
Cascaded Model	0.83	1.03	0.55	0.82

The relative importance, or strength of each module can be determined through examining Table 4. While the cascaded model performed the best overall, with a correlation of 0.83, several individual modules stand out. The three strongest independent modules in regards to improved correlation were the spatial localization, as well as the two spatial frequency adaptation modules. It is interesting to note that the image dependent adaptation module and the spatial localization module both predicted the less sharp images very well, while sacrificing the performance in prediction of the sharper images. The natural scene adaptation (1/f) improved upon the less sharp images slightly less, but also improved prediction of the sharper images. This analysis describes the improvement in overall correlation each of the modules provide, but does not provide much insight on whether the image difference module can be used as a replacement for actual psychophysics. An analysis of this nature is discussed further below. The correlation analysis does provide empirical insight into the relative strength of an image difference model, and the gain of any of the individual modules.

All of the modules predicted an intercept greater than 0, with the cascaded full model showing the smallest intercepts of 1.03 and 0.82 for the negative and positive predictions respectively. This indicates that there is a relatively large spread in predicted differences that were judged to be identical in sharpness. This spread represents the range of error that was either below threshold, or did not alter the perception of

sharpness. The extent of this spread might be an artifact of the spatial filtering in the first stage of all the modules. The contrast sensitivity functions used were all normalized to be 1.0 at the DC component, resulting in values greater than 1.0 for certain frequencies. This has the effect of modulating and enhancing certain frequencies. Perhaps it is this enhancement that is causing even slight physically measurable errors to be boosted, resulting in an intercept greater than 0. It might be interesting to see if this type of filter proves useful for threshold models.

For an ideal model, it can be argued that the slope on both sides of the “V” should be identical, and also that the intercept must be forced to 0. This type of model suggests that “good” differences and “bad” differences are perceptually similar. Conceptually this might be ideal, but in practice this might not be the case. Consider the effect of contrast on the perception of sharpness. Increasing contrast (by changing the slope of the lightness channel) might serve to increase the perception of sharpness up to a point, and then might actually serve to decrease the perception of sharpness (perhaps due to clipping or quantization). An image difference model might predict a larger difference caused by the large contrast change even if it has a detrimental effect on the perception of sharpness. This can be thought of as “looping back” through the perception of sharpness. This would lead to an increased slope for the positive side. For the negative side, a decrease in contrast might be linearly related to a decrease in the perception of sharpness. As the contrast decreases, the image difference increases. Generally this decrease will continue, and not “loop-back” to a perceptual increase in sharpness. Thus the negative perception of sharpness might involve a constant linear slope, while the positive perception might have a different slope.

It is important to determine whether this is the case, or whether the slopes of the positive and negative should be identical. Analysis of this nature can be accomplished by fitting a single regression line to the absolute value of the normalized interval scale. This analysis can be further constrained by forcing this single regression line to have a 0 intercept. It should be noted that by forcing this constraint the regression analysis is no longer a least-squares solution. The correlation for a non-least squares solution is not represented by the square of the Pearson coefficient, but is rather representative of the model sum-of-squares divided by the residual error sum-of-squares. Further discussion of this relationship is given below.

Figure 110. Sharpness Predictions Using Identical Slopes and 0 Intercept

Table 5. Correlation Coefficients for Identical Slope

Model	No Intercept	Intercept
Enhanced Movshon CSF	-1.44	0.32
Spatial Localization	-0.03	0.75
Local Contrast	-1.35	0.23
1/f Spatial Adaptation	-0.81	0.59
Image Dependent Adaptation	0.39	0.86
Cascaded Model	0.39	0.78

This analysis provides interesting insight into the model behavior, as illustrated by the correlation coefficients shown in Table 5. When the intercept is forced to zero the correlation coefficient actually goes negative for several of the modules. This indicates a weakness in the correlation metric for this particular regression, caused by an inability to minimize the least squares error. This error is often the result of an incorrect choice of algorithm.^{77,78,79} The negative correlation actually suggests that the average model prediction across all images is better at predicting the experimental data than any individual prediction. The correlation for this non-least squares solution is calculated by using the Equation below.

$$R^2 = \frac{MSS}{TSS} \quad (36)$$

where MSS is the model sum of squares, and TSS equals the total sum-of-squares. The model sum-of-squares is calculated as follows:

$$\begin{aligned} MSS &= TSS - RSS \\ TSS &= \sum (y - \bar{y})^2 \\ RSS &= \sum (y - xb)^2 \end{aligned} \quad (37)$$

where y represents the value of any given point, $\bar{y}$ represents the mean value of y , x is the regressor variable, and b is the slope of the fit line. For non-least squares solution the residual error is not constrained to be lower than the total sum-of-squares. Forcing the intercept to zero can allow the residual errors to be larger than the total sum-of-squares, indicating that $\bar{y}$ is a better predictor (has smaller error) than the modeled fit. Essentially the negative correlation implies a very poor statistical fit, or a poor choice of regression models. In this situation it implies the importance of the intercept for the least-squares solution. For the image dependent spatial frequency adaptation, as well as the cascaded model,

the non-least squares regression was able to find a solution that did not have a negative correlation, thus suggesting a suitable statistical fit.

When the regression model is given the freedom to add an intercept, or offset term, all of the models produce a positive correlation coefficient. In this situation the image dependent adaptation metric proved to be a more accurate predictor of the experimental data than the cascaded model (0.86 to 0.78 respectively). This indicates that one of the modules actually decreases the predictive capability. By examining the plots in Figure 109 and Figure 110 it is obvious that the local-contrast module predicts a single image to be of much larger difference than the experimental results suggest. This difference is evidenced in the positive sharpness direction, indicating that perhaps there is a situation where increasing contrast resulted in a decrease in sharpness (too much contrast causing a loop-back). The contrast module could also be simply amplifying error or noise in its prediction.

This type of analysis reveals the importance of having the intercept term in the model prediction. This suggests a spread in images with physically measurable differences that are not perceived to be different (or at least to be of different sharpness). This spread either suggests an over-prediction of error around threshold differences, or could also represent the threshold difference spread (equivalent to the ellipses in traditional color difference formulae). If this is an over-prediction of error that is below threshold, then perhaps a visual masking module designed to suppress model output near threshold could reduce the need for an intercept. Just as research into small color differences might suggest a threshold of 1.0 CIE ΔE^*_{ab} an experiment could be designed to find the perceptibility threshold of the image difference metric and determine if the intercept term is below that value.

9.1.10 Further Model Analysis: Replacing Psychophysics?

The empirical metrics described above give insight into the relative strengths of the model predictions, and allow us to compare different models with each other. This in-and-of itself is a useful tool, as it provides an important tool for comparing and developing image difference models. What these analyses do not provide is a method for determining whether these models are capable of replacing or augmenting psychophysics. Further analysis that takes into account observer variability and calculates “false” positives and negatives can potentially provide more insight into these models usefulness.

An example of this type of analysis is provided below. The experimental variability first needs to be taken into account, as it is not possible for a model to be better than actual human observers. An analysis can be performed by examining the model predictions using the identical slope method, such as Figure 110, with the addition of lines corresponding to the observer variability. These lines are calculated by taking the equation of the linear trendline in conjunction with the confidence intervals from the experiment. The equation for the observer variability lines then becomes:

$$y_{\text{var}} = mx + b \pm CI \cdot m \quad (38)$$

where m is the slope of the best-fit regression line, b is the slope of the regression line, and CI is the confidence intervals. An example of this type of analysis for the Movshon frequency enhancing CSF with 90% confidence intervals is shown in Figure 111.

Figure 111. CSF Model Predictions with Observer Variability

It is then possible to count how many of the predictions lie inside or on the observer variability lines (solid) and how many are above and below. The observations that lie below the lines indicate false negatives, or images that are perceived to be sharper than the model predicts. The observations above the lines indicate false positives, or images that were perceived to be less sharp than the model predicts. Using just the CSF filters the model predicts 18 images correctly, 25 false negatives, and 28 false positives. This indicates that the image difference model using just CSF filtering alone is not adequate for the replacement or augmentation of psychophysical experimentation. This analysis can be performed for all the various modules, to determine if they are viable alternatives to full-scale experimentation. It is also important to have an idea as to what errors are more acceptable, false negatives or positives. If the goal is to augment psychophysics, such as using the image difference metric to eliminate obvious images from a large paired-comparison experiment, than it might be tolerable to have false negatives. These would be images that could be excluded from an experiment but are included anyway. While this would result in more observations, the experiment itself would produce the correct scale. Likewise it might be more acceptable to have false positives than false negatives, especially when dealing with threshold data. For instance in medical imaging an application might be examining current images with known “healthy” originals. A false positive would mean that a radiologist might have to examine a healthy image manually that was flagged automatically. A false negative, on the other hand, would essentially mean that an image

that a radiologist should see would never be flagged. It should be stressed that this is just an example, and not a targeted (nor necessarily endorsed) application.

This type of experimental variability analysis is shown in Figure 112 and summarized in Table 6.

Figure 112. Module Predictions Accounting for Observer Variability

Table 6. Experimental Predictions with Observer Variability

Module	False Negative	False Positive	Correct	Percent Correct
CSF Filter	25	28	18	25%
Spatial Localization	10	9	52	73
Local Contrast	25	25	21	30%
Image Dependent Adaptation	7	8	56	79%
Cascaded Model	11	12	48	68%

This analysis shows that some of the modules predict less than half of the actual observations correctly. This includes using just the CSF alone (such as the traditional S-CIELAB) and using it in conjunction with the local-contrast module. The spatial localization module and the image dependent spatial-

frequency adaptation module provide the best predictions, at 73% and 79% respectively. It is interesting to note that the cascaded module predicts on 68% of the observations correctly. This indicates that one of the modules is actually hurting the model predictions. This analysis indicates that the local-contrast module might be the cause of this decrease in performance.

By taking into account the observer variability, as well as the models ability to predict experimental results we gain a better understanding in how we can apply these models in the future. For example the model utilizing the CSF filtering in conjunction with image dependent adaptation was capable of predicting the experimental results within 80% of the human variability. It would be possible to use this model to augment and aid psychophysical experimental design. For instance the sharpness experiment could be redesigned with a smaller population of images by choosing an image difference threshold, and throwing away images that fall above that threshold. Knowledge of the model performance can help choose that threshold.

9.1.11 Sharpness Experiment Conclusions

The sharpness experiment described in this section has provided a wealth of data with which to evaluate the performance of the color image difference metric and the individual modules that make up the metric. Before we can test these predictions with empirical metrics we must first qualitatively understand if they are behaving as expected. This is important, as it was previous unknown as to whether we could use an image difference metric to predict perceptual scales of sharpness. Each of the modules described above was shown to behave in a manner that was expected and followed the general trend expected. This qualitative analysis can provide insight into the model behavior, but it cannot provide insight into the relative performance gain of each of the modules. Several empirical methods were examined to establish the correlation between model output and experimental sharpness scales. Using this analysis the individual modules could be evaluated. In general each module showed improvement in the experimental predictions on its own, while cascading the modules together was the most accurate at predicting the experimental results when the negative and positive sharpness scales were examined separately, as illustrated by Figure 109 and Table 4. When forcing the slopes to be identical, we see that the local-contrast module over-predicts certain image differences, as shown in Figure 110. Furthering this analysis by taking into account observer variability it is possible to determine the number of correct experimental predictions, as well as the number of false positives and negatives. This analysis indicated that the image-dependent spatial frequency adaptation performed the best.

9.2 Contrast Experiment

The data from the Contrast Experiment can be analyzed in identical ways as the data from the Sharpness Experiment. The interval scales created in the Contrast Experiment can also be thought of as difference scales, where the difference refers to the change in perceived contrast. Once again, these scales are normalized so that the “original” image has a perceived contrast of “0.” Images that have a positive scale value are judged to be higher in contrast than the original, while images with negative scale values are judged to be lower in contrast.

The manipulated images are used as input into the image difference framework along with the single original image. Thus the perceived difference between all the image manipulations and the original is calculated. If the image difference metric correlates with the experimental data, we would expect to see the same “V” shaped curve as described in the results of Sharpness Experiment. It is important to note again that the image difference has no mechanism for determining whether a perceived contrast difference is an increase or a decrease.

9.2.1 Lightness Experiment

The mean value of the error image using the Movshon CSF, modified to be anisotropic with image dependent spatial frequency adaptation is shown in Figure 113 plotted against the experimental contrast scale. The experimental scale is averaged over all image scenes.

Figure 113. Model Predictions of Lightness Experiment Contrast Scale

The model is shown to very accurately predict the experimental data, as evidenced by the tight “V” shaped distribution. The correlation coefficient for the image deemed to be of less contrast is an impressive 0.98, while the correlation for the higher contrast images is 0.74. The intercepts for both series are also above 0, at 0.61 and 0.57. This again suggests that the frequency enhancement of the contrast

sensitivity function boosts even small errors. From Figure 113 we can see that a simple image difference model using just the spatial filtering and adaptation is capable of predicting this data-set. The univariate nature of the contrast experiment does not have the inherent interactions between several cascaded manipulations, making it ideal to test the image difference framework. Since the observers were scaling a perception based upon a single dimension of manipulation the relationship between that perception scale and the overall image difference should be related. Further model analysis that forces the slopes to be identical and also forces the intercept to zero can also be performed. This is illustrated in Figure 114.

Figure 114. Model Predictions of Lightness Contrast Scale with Identical Slope

The metric predictions with a 0 intercept has a correlation coefficient of 0.70. When an intercept is allowed for, the correlation increases to 0.78. This decrease in performance when identical slopes are forced suggests that the relationship between increased contrast and perceived difference is not necessarily linear. Perhaps when the positive image difference increases too much, the images begin to decrease in contrast (loop-back effect).

9.2.2 Chroma Experiment

The results of the same image difference metric are shown below plotted against the experimental contrast scale created from the chroma experiment. Again, these calculations are performed using the Movshon CSF filtering with the image-dependent spatial frequency adaptation model. The means of the error images are plotted with the experimental contrast scale in Figure 115.

Figure 115. Model Predictions vs. Perceived Chroma Contrast Scale

The image difference predictions are highly correlated with the contrast scale, with the exception of a single data point. The correlation coefficient for the images judged to be of less contrast is 0.90. As there are only two data points for the images judged to have more contrast, it is not necessary to calculate the correlation (automatically 1.0). The experimental results show an almost monotonic increase in contrast with chroma, except for when there is very little chroma added to an image. The image with no chroma (grayscale) was judged to have significantly higher contrast than the image with 20% chroma. There are several theories as to the cause of this perception expressed by Calabria.⁷³ The image difference model is not capable of making this distinction. Forcing identical slopes for the increasing and decreasing contrast images does not change the model prediction much, as shown in right side of Figure 115. There is not a nonlinear looping relationship with excessive chroma boosting like there was with lightness. Perhaps this is because the chroma boosting was limited to a 1.2 increase of the original chroma.

9.2.3 Sharpness Experiment

In the sharpness experiment all of the manipulations were judged to have more contrast than the original image, so the model predictions do not have the characteristic “V” shaped trend. Rather there is a single linear trend. The mean image difference predictions using the Movshon CSF and image-dependent spatial frequency adaptation are shown in Figure 116.

Figure 116. Model Prediction vs. Sharpness Contrast Scale

The image difference framework was able to very accurately predict the results of the monotonic increase in contrast caused by the unsharp mask, with a correlation coefficient of 0.96. Forcing the slope to 0 results in little change, with a correlation coefficient of 0.95.

9.2.4 Psychophysical Replacement?

Analysis of the contrast experiment in conjunction with the observer variability can be performed, similarly to Section 9.1.10. From this analysis it is possible to calculate the number of correct predictions, as well as the number of false negatives and positives. Figure 117 shows the contrast experiment predictions along with the confidence intervals representing observer variability. The image difference metric predicted 3 false positives for the lightness experiment, 1 false positive for the chroma experiment, and correctly predicted the sharpness experiment. The three false positives for the lightness experiment corresponded to three images that were judged to have more contrast than the original. In these cases the model predicted a larger difference than the observers did. This could be a result of over prediction, or it could be that the lightness manipulations caused a larger perceived difference that did not correspond to an increase in contrast. The single false positive for the chroma experiment corresponded to the achromatic image. This image was judged to have more contrast than the image with 20% chroma, though the image difference model saw it to be the most different.

Figure 117. Contrast Experiment Predictions with Observer Variability

9.2.5 Contrast Experiment Conclusions

The experimental results from the Contrast Experiment were highly correlated with the image difference predictions using just the spatial filtering and spatial adaptation modules of the image difference framework. This is encouraging, as the univariate nature of the experiment proved to be a very good test of simple image differences. The modular nature of the framework allowed for a choice of a relatively simple model in comparison to the full cascaded model described in Section 9.1.6, as that proved sufficient for predicting this data. This follows with the guidelines of keeping the model only as complicated as necessary. When the observer variations were taken into account, the model performance was very good, with 100% success for the sharpness manipulation, and 85% for the chroma and lightness manipulations. This suggests that an image difference metric could be used successfully to augment psychophysical examination for a single varying image manipulation.

9.3 Print Experiment Predictions

There were three separate rank-order print experiments, corresponding to perceived sharpness, graininess and overall image quality. This section outlines the image difference model predictions for these three

experiments. It should be noted that the model predictions for all three experiments are identical, as the perceived image differences did not change between the image pairs. Thus this experimental dataset can lend insight into the relationship between image differences and three distinct perceptions, or “nesses.”

9.3.1 Sharpness Experiment

In this experiment observers were asked to rank the images in the order of sharpness. The experimental results for this particular experiment did not match up well between the RIT and the Fuji datasets, indicating a difficulty in this judgment or a discrepancy between the groups of observers. The Fuji experiment was performed entirely by male experts, while the RIT experiment included a mixture of male and female, expert and naïve. The data is predicted using the image difference modules of the Movshon anisotropic CSF, along with image dependent spatial adaptation and with spatial localization. The mean experimental predictions for the Ship image for both the Fuji and RIT data are shown in Figure 118.

Figure 118. Image Difference Predictions of Sharpness (Ship)

The model prediction of the images that are perceived to be sharper is more correlated than those predicted to be less sharp, with a 0.87 and 0.55 correlation coefficient for the Fuji and RIT data respectively. The model does not do a good job predicting the image deemed to be less sharp, evidenced by correlation coefficients of less than 0.1 for both datasets. The predictions are substantially worse for the Fuji ranking of the portrait image, as illustrated in Figure 119.

Figure 119. Image Difference Prediction of Sharpness (Portrait)

The model shows a stronger correlation for the RIT data than the Fuji data, though both are below 0.5. The RIT correlation is pulled heavily from the origin to a cloud of observations, indicating that the statistical weight of the origin is too much (outlying). Overall the model seems incapable of predicting the Fuji data. The trend line actually goes in the opposite direction for the images deemed to be less sharp. The RIT and Fuji data did not match well for this particular scale, as described in Section 8.3, and Figure 73. This indicates that the observers struggled with this attribute, perhaps as a result of interjecting some preference into the perception of sharpness in portrait images. The discrepancies between the RIT and Fuji data might also result from the use of a homogenous observer group at Fuji (all male and experienced), while the RIT result was a mix of male, female, experienced and naïve. The Fuji group could have used different criteria in their judgment of sharpness than the RIT group. The inability of the model to predict this particular scale also suggests attribute interactions between the various interactions. It should be noted that there are two distinct groups of images predicted from the image differences (below 10, and above 10). These correspond to the two ISO speed differences, as the ISO 300 image with no manipulations was used as the original. It is interesting to note that the RIT group deemed all the ISO 300 manipulations sharper than the corresponding ISO 1600 images, while the Fuji group did not have this same distinction. This supports the notion that the Fuji group used different criteria in their judgment of sharpness.

Forcing identical slopes for all predictions, as well as eliminating the intercept, results in similarly bad predictions as seen in Figure 120. This indicates that perhaps the perception of sharpness for this experiment is not related to the perception of overall image difference. The image difference metric was capable in distinguishing the difference between the ISO speeds. This information might be useful in designing new (smaller) experiments that rank sharpness separately between the two speed levels. It is not useful to calculate the number of false positives and negatives for this particular experiment, as they certainly outnumber the correct predictions.

Figure 120. Sharpness Prediction with Identical Slope and 0 Intercept

9.3.2 Graininess Prediction

In the graininess experiment observers were asked to rank the images based on the perception of graininess. The mean image difference predictions using the same modules, with the ISO 300 image as the original are shown in Figure 121 for the Ship image.

Figure 121. Image Difference Predictions of Graininess (Ship)

The image difference model shows a strongly correlated relationship with the Fuji data, with a correlation coefficient around 0.95. The Fuji group determined all the images were grainier than the original, while The RIT group predicted the original to be grainier than several of the other manipulations. The image difference predictions also are strongly correlated with the RIT data, though less so than the Fuji data. Perhaps this is because the RIT observers are not as experienced in this type of observation, as the Fuji engineers are. Forcing the slopes to be identical as well as the intercept to be 0 results in the predictions shown in Figure 123.

Figure 122. Image Difference Predictions of Graininess with Identical Slopes and 0 Intercept

Forcing the slopes identical actually improves the RIT prediction, increasing the correlation coefficient to 0.91 when allowing for an intercept. Both datasets show a slight decrease in performance when eliminating the offset term. This again suggests a slight spread in model prediction for images at or near threshold. The model predictions for the portrait image are shown in Figure 104.

Figure 123. Image Difference Predictions of Graininess (Portrait)

The image difference model predictions for the portrait image correlate very well with the experimental data, with coefficients greater than 0.9 for the images judged grainier. Forcing identical slopes, and eliminating the intercept results in the predictions shown in Figure 124.

Figure 124. Image Difference Predictions of Graininess with Identical Slopes and 0 Intercept

Forcing the slopes to be the same actually increases the prediction slightly for both groups, suggesting a linear relationship between the model predictions and the perception of both increasing and decreasing graininess. Removing the intercept term decreases the correlation slightly, again suggesting a possible threshold range for the image difference predictions.

Taking into account the observer variability shows no false negatives or positives for the RIT data, and 3 false positives and 2 false negatives for the Fuji data. This relationship is shown in Figure 125. The RIT data did not indicate significant differences between the images inside each ISO speed group, which is why the predictions all fall within the variability range.

Figure 125. Graininess Predictions with Observer Variability

That the same image difference predictions correlate very well to the graininess scale and not the sharpness scale indicates that perhaps graininess is a simple image difference perception, while sharpness might be a higher order perception.

9.3.3 Image Quality Experiment

Up to this point the image difference framework has been used to predict various percepts, or “nesses,” such as sharpness, contrast, and graininess. The print experiment offers the first test for predicting overall image quality. It is important to realize that the judgment of quality in the experiment is most likely influenced by the changes in graininess and sharpness in the image manipulations, as well as from the previous ranking experiments. Figure 126 shows the mean image difference predictions of the image quality experimental scale for the Ship image.

Figure 126. Image Difference Predictions of Quality (Ship)

The relationship between the image difference model predictions and the experimental quality scale show correlation coefficients of 0.77 and 0.86 for the Fuji and RIT images judged to be lower quality than the original. Both the RIT and Fuji datasets are overwhelmingly influenced by the ISO speed, as all of the ISO 1600 images were deemed to be of much lower quality than the ISO 300 images. Figure 127 shows the predictions by forcing the slopes to be identical for both higher and lower quality images, as well as removing the intercept term.

Figure 127. Image Difference Predictions of Quality with Identical Slopes and 0 intercept

The correlations for the Fuji data are slightly worse when the slopes are identical and an intercept is allowed. These predictions get worse yet when the intercept is forced to 0. The RIT data is well predicted

with identical slopes, with a correlation of 0.86 and 0.79 with and without an intercept respectively. The predictions for the portrait image are shown in Figure 128.

Figure 128. Image Difference Predictions of Quality (portrait)

Figure 108 shows very strong correlations between the image difference metric and perceived image quality for the portrait image. This relationship is strong for both the RIT and Fuji data, and equally strong when the slope of the predictions are identical for increases and decreases in perceived quality.

That the image difference model correlates well with image quality scales indicates the potential use of this type of model for building an overall image quality metric. In this particular experiment the quality scale was highly correlated with the graininess scale. Because the image difference metric was able to predict the graininess scale, it was capable of predicting the quality scale. This notion of image quality modeling is explored further in Section 10.

9.3.4 Print Experiment Summary

Three hard-copy ranking experiments were used to evaluate the performance of an image difference model based on the modular framework. Of the three experiments, the image difference metric proved quite capable of predicting the perceived graininess scales, as well as overall image quality. The model

struggled with predicting the sharpness scale. From the discrepancies between the RIT and Fuji datasets it seems that the observers themselves also struggled with scaling sharpness, especially for the Portrait image. The model predicted large discrepancies between the ISO speed images, which played a large part in the perception of all the attributes. This suggests one potential use for the model already, in the designing of the psychophysical experiments. These experiments might have been better served by first calculating the image differences. The large differences in speed indicate the need to have these two groups separated into two distinct experiments. This would allow for a more in depth analysis of the other manipulations inside the ISO speed groups.

9.4 Psychophysical Experimentation Summary

Section 9 has detailed the use of several psychophysical datasets in the design and evaluation of the color image difference framework. Section 9.1 outlined a softcopy paired comparison experiments examining the perception of sharpness. This dataset has been used to evaluate the performance of all the individual modules in the framework, as well as the cascaded model performance. A qualitative analysis was performed to establish the general behavior of the modules, and to determine if an image difference metric was capable of predicting a perceptual attribute such as sharpness. This analysis was extended to include quantitative metrics examining the correlation between the model predictions and experimental sharpness scale. Each module was shown to increase the model prediction of the data, though some modules were more beneficial than others. The summary of each of the modules relative performance can be seen in Table 4. This analysis was extended further to take into account the inherent observer variability. From this analysis it is possible to determine the number correct predictions, as well as the number of false positives and negatives. Using this type of analysis it was determined that the image-dependent spatial frequency adaptation along with the CSF filtering performed the best, predicting roughly 80% of the correct responses. This suggests that the image difference metric could be a viable tool for augmenting psychophysical experiments.

Two other independent datasets, created from a series of soft-copy and hard-copy experiments, were used to test the performance of the image difference model. The contrast experiments scaled perceived contrast across a series of individual manipulations. The image difference metric, created by cascading the modules together, proved to be very successful in predicting the results of the perceived contrast resulting from changes in lightness, chroma, and sharpness. The prediction success was represented by the high correlations as well as a low number of false predictions. The print experiment scaled sharpness, graininess, and overall quality as a result of a series of image manipulations typically found in the design of digital cameras. The image difference metric was capable of predicting the graininess and overall image quality scales well, though it struggled with the sharpness scale. The image

quality scale for this experiment was highly-correlated with the graininess scale, which might explain why the image differences were able to predict the perception of quality. The sharpness scale, on the other hand, could not be predicted well at all. This scale was also the most different between the RIT and Fuji groups, indicating, perhaps, differing judgment criteria between the two groups. The unique properties of the sharpness scale could be a result of the multi-variate nature of the sharpness perception, or a result of the observer noise itself.

The image difference metric is incapable of determining causes or direction of perceived differences, only magnitudes of difference. Section 10 discusses techniques that can be used to begin to understand the cause, and direction, of differences.

10 Image Appearance Attributes

This section describes the evolution of the color image difference framework towards a model of image appearance. An image appearance model can be thought of as a color appearance model for complex spatial stimuli. This allows for the prediction of color appearance attributes such as lightness, chroma, and hue, as well as spatial 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.

Recall that an image difference model is only capable of predicting magnitudes of errors, and not direction. A model capable of predicting perceived color difference between complex image stimuli is a useful tool, but it does have 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.

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 is analogous to a traditional color difference equation such as CIE ΔE^*_{ab} . Traditional color difference equations only tell the magnitude of the perceived difference, and not the direction or cause of said difference. This information can be obtained by examining the individual color changes themselves, such as ΔL^* , ΔC^* , and Δh . These changes are not Euclidean distances so they maintain direction and magnitude information. Thus it is possible to determine the cause of an overall color difference, such as a hue rotation. The modular framework allows for similar calculations by examining the output from each of the individual modules. This can be beneficial for determining the root cause of an overall error image. For instance, if someone is designing an image reproduction system she can use the overall image difference metric to determine the perceived magnitude of error. She can then examine the output from the individual modules to determine the cause of the error, such as change in contrast. Figure 129 illustrates this principle.

Figure 129. Using Individual Modules for Determining Cause of Image Difference

This concept can be explored using the Sharpness Experiment dataset, as there were several simultaneous image manipulations.

10.1 Resolution Detection

As described in Section 8 there were three levels of resolution, or addressability, tested in the Sharpness Experiment corresponding to 300, 150, and 100 pixels-per-inch. The spatial filtering module, in conjunction with the spatial frequency adaptation, should be able to detect these three levels of resolution differences when compared against the original 300 dpi image. This is accomplished by examining the standard deviation of the ΔL^* channel output from this module. The L^* channel is used because of the nature of the spatial filters, as the luminance channel is much more sensitive to high frequency differences. The standard deviation of this channel is thought to best detect the changes in resolution, as the error image will have relatively small error in the low frequencies, and much larger errors in the high frequencies where the low resolution images contain no information. This combination of small and larger errors results in a large standard deviation. The prediction of the standard deviation of the L^* channel of the sharpness scale data is shown in Figure 130.

Figure 130. Standard Deviation of L^* Channel

There are three relatively distinct groupings shown in Figure 130, highlighted by the red ovals. These groups correspond to the three levels of resolution. The three groups are not completely separated as indicated by the overlap between two of the ovals and the differing orientations. This overlap is most likely caused by one of the other image manipulations adding high-frequency information. Thus this metric itself is not fully capable of detecting resolution changes in a fully automated manner. The strength of this type of analysis is that it would allow a color researcher to examine one of the image difference plots shown in Section 9.1, along with this type of plot, to determine if the cause of the image difference was a change in resolution.

10.2 Spatial Filtering

Spatial filtering using a convolution Laplacian edge enhancement was applied to half of the images in the Sharpness Experiment. Examining the output of the spatial localization module should be able to detect whether spatial filtering was applied or not. The spatial localization module is typically applied to the luminance opponent channel, by filtering with a Gaussian kernel. We can examine the standard deviation of the difference of the two luminance channel images. In Figure 131 the luminance images are filtered with a Gaussian centered at 20 cycles-per-degree with a width of 5 cpd.

Figure 131. Prediction of Spatial Filtering

There are two distinct groups shown in Figure 131, corresponding to the two levels of spatial filtering applied to the images. This type of plot can be useful to a research as it also reveals more information into the causes of the perceptual sharpness scale. The resolution prediction, shown in Figure 130, appears very similar to the general “V” trend of the typical image difference predictions. The three resolution levels are fairly distinct along the sharpness scale (x-axis) indicating the importance that resolution played a key role in the perception of sharpness. Figure 131 does not have this distinctive separation, as the two levels of spatial filtering span the entire sharpness scale. This indicates the smaller role that spatial filtering had on overall perceived sharpness when compared to resolution, as discussed in Section 8.1.6. An analysis of the localization module, as shown in Figure 131, could provide a good indication that an image difference is caused by spatial sharpening.

10.3 Contrast Changes

There were three levels of contrast in the Sharpness experiment. The local-contrast module was designed to detect changes in contrast, so it stands that the output of that module should detect these three levels of contrast. 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 the mean difference of the CIELAB L* channel output from the contrast module can be plotted. This is shown in Figure 132.

Figure 132. Prediction of Changes in Contrast

It is clear that there are three distinct groups in the Figure 132, corresponding to the three levels of contrast in the input images. A researcher could examine an overall image difference map, along with this type of plot to determine if the cause of the error was a change in contrast. Notice once again that each level of contrast spans the entire sharpness scale. This indicates that contrast, while playing an important role in perceived sharpness, was overshadowed by other manipulations such as resolution.

10.4 Putting it Together: Multivariate Image Quality

This section has outline a potential use of the modular image difference framework for pulling out information relating to the cause of perceived image differences. This can be useful for building multivariate models of image quality, using similar techniques as described by Keelan.⁶ One simple technique for image quality scaling is through the weighted sum of various perceptual attributes such as contrast and sharpness. We can use the same type of techniques, not for overall quality scaling but instead

for scaling a single perception such as sharpness. Up to this point the image difference metric was capable only of predicting magnitude, and not direction, of sharpness. By combining the output of the individual modules it should be possible to generate a prediction of sharpness direction. The output of the contrast module, the spatial localization module, and the overall image difference error map has been linearly combined (after normalization to similar scales) as an ad-hoc magnitude and direction metric. The predictions of this type of metric are shown in Figure 133.

Figure 133. Prediction of Ad-hoc Magnitude and Directional Metric

This prediction is presented as a “proof-of-concept” rather than as a complete attempt to predict the overall experimental scale. What is interesting is this type of modeling can start to predict both magnitude and direction of the experimental sharpness scale. The arrow shows the general trend in Figure 133 indicating the direction of increased sharpness. This type of metric could then be combined with other image appearance attributes, such as contrast and graininess, to create an overall image quality model.

10.5 Image Attribute Summary

This section has outlined steps that can be taken to predict root causes of image differences. One of the strengths of the modular framework is that it allows for output and examination of image data at each individual module. This output can be used to provide additional information to an end-user, and can help determine what types of image difference are encountered. The output from the various modules was shown to detect the changes in resolution, contrast, and spatial filtering of the input images. By examining the individual output of the module it is also possible to determine both the magnitude and direction of the image differences. This is similar in nature to examining the individual components of a color

difference formula, to determine the cause of the color difference. For example, the overall CIE ΔE_{94} might provide the magnitude of color difference, while the individual ΔL , ΔC , and ΔH components describe the direction of the difference.

In the current state the module outputs have no real perceptual meaning. For instance, the output of the local contrast metric was capable of detecting changes in “gamma” of the input images, but it cannot be said that the output is a metric of the appearance of perceived contrast. More information is necessary to determine these perceptual scales. Just as research in color appearance has had many years scaling perceptual attributes such as lightness, chroma, brightness and hue to develop CIECAM02, it might take as many years to develop scales of spatial appearance correlates such as sharpness, graininess, and contrast. A framework for this type of model, known as an image appearance model, is presented in Section 11.

11 ICAM: an Image Appearance Model

This dissertation has outlined the motivation for, and the inspiration behind the creation of a modular color image difference framework. The concept behind this framework has been discussed along with many individual modules that when combined create a metric that is very capable of predicting perceived image differences.

At the heart of the modular framework lies the “core metric.” This metric is a color space, and has been CIELAB along with the CIE color difference equations throughout most of the discussion in the previous sections. If the core metric were to be replaced with an appearance space, then an image appearance model is born. This image appearance model shares the same strengths as the image difference framework, namely simplicity and modularity. The foundation for such an image appearance model has been laid, and has resulted in the formulation of iCAM, the Image Color Appearance Model.^{80,81}

The image difference model was design to augment traditional colorimetry and color difference formulae to account for spatially complex stimuli. An image appearance model is similarly designed to augment and extend traditional color appearance modeling with aspects of spatial vision. Color appearance models themselves extend upon traditional colorimetry to include the ability to predict perceptions of colors across disparate viewing conditions. Color appearance models map physical measurements into appearance correlates such as lightness, chroma, and hue. When applied to digital images, color appearance models have traditionally treated each pixel in the image as an independent stimulus. Image appearance models attempt to combine the color appearance aspects with the spatial properties of the human visual system. The goal of image appearance modeling is to extend the prediction of color correlates to include spatially complex correlates such as sharpness, graininess, and contrast.

At the heart of iCAM lies the IPT color space, as described in Section 7.6.1. This space serves as the core metric, and is augmented with several of the ideas described in this dissertation as the modular framework for color image difference calculations. These modules are extended to include spatial models of chromatic adaptation, luminance adaptation, and viewing surround. The general flowchart for spatial iCAM is shown in Figure 134.

Figure 134. Flowchart for iCAM: a Spatial Image Appearance Model

The goal in the formulation of iCAM is to combine the research of color appearance, spatial vision, and color difference into a single unified model. This type of model is applicable to a wide range of situations,

including, but not limited to, high-dynamic range imaging, image color difference calculation, and spatial vision phenomena.

The input into the iCAM model, as shown in the top of Figure 114, is a colorimetric (CIE XYZ) image. The adapting stimulus, or “whitepoint” is specified as low-pass filtered version of the input image. The adapting image can also be tagged with absolute luminance information, which is necessary to predict the degree of chromatic adaptation. This absolute luminance information of the image is also used as a second low-pass image to control various luminance-dependant aspects of the model. These are necessary to predict the Hunt effect (increase in perceived colorfulness with luminance) and the Stevens effect (increase in perceived image contrast with luminance).⁸¹ Finally another absolute luminance input image of larger extent is necessary to predict certain surround effects.

The image and the adapting “white” image are first processed through a von Kries type chromatic adaptation transform. The chromatic adaptation transform comes from the recently-published CIECAM02 model⁸², and has been adopted in iCAM since it was well researched and established to have excellent performance with all available visual data. It is also a relatively simple chromatic adaptation model amenable to image-processing applications. The specific chromatic adaptation model, given in Equations 39-44, is a linear von Kries normalization of sharpened RGB image signals by the RGB adaptation signals derived from the low-pass adaptation image at each pixel location ($R_w G_w B_w$).

$$\begin{bmatrix} R \\ G \\ B \end{bmatrix} = \mathbf{M}_{CAT02} \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} \quad (39)$$

$$\mathbf{M}_{CAT02} = \begin{bmatrix} 0.7328 & 0.4296 & -0.1624 \\ -0.7036 & 1.6975 & 0.0061 \\ 0.0030 & 0.0136 & 0.9834 \end{bmatrix} \quad (40)$$

$$D = F \left[1 - \left(\frac{1}{3.6} \right) e^{\left(\frac{-L_A - 42}{92} \right)} \right] \quad (41)$$

$$R_c = \left[\left(Y_w \frac{D}{R_w} \right) + (1 - D) \right] R \quad (42)$$

$$G_c = \left[\left(Y_w \frac{D}{G_w} \right) + (1 - D) \right] G \quad (43)$$

$$B_c = \left[\left(Y_w \frac{D}{B_w} \right) + (1 - D) \right] B \quad (44)$$

The RGB signals are computed using a linear transformation from XYZ to RGB derived by CIE TC8-01 in the formulation of CIECAM02. This matrix transformation has come to be called the MCAT02 matrix, where CAT stands for chromatic adaptation transform. The input image, and the low-passed “white-point” image are both converted into sharpened RGB signals using Equation 39. The von Kries normalization is further modulated with a degree-of-adaptation factor, D , that can vary from 0.0 for no adaptation to 1.0 for complete chromatic adaptation. Equation 40 is also from the CIECAM02 formulation, and is used in iCAM for computation of D (degree of adaptation) as a function of adapting luminance, L_A , for various viewing conditions. Alternatively the D factor can be established manually. The chromatic adaptation model is used to compute corresponding colors for CIE Illuminant D65 that are then used in the later stages of the iCAM model.

The form of this transform is identical to that used in CIECAM02,⁶⁵ except that the adapting white is spatially variable. Essentially this means that the images are adapted based on their local neighborhood, rather than using a single “gray-world” number. The extent of the adaptation neighborhood (amount of low-pass blurring) depends upon application of the model.

After adaptation the image is transformed into the IPT color space, using the equations described in 7.6.1. One important variation from the traditional IPT space is the use of a spatially modulated exponent value combined with the 0.43 exponent in Equation 31. This modulation is controlled by the F_L function from CIECAM02 given in Equation 45. This function, which is slowly varying with luminance, has been established to predict a variety of luminance-dependent appearance effects in CIECAM02 and earlier models. Since the function has been established and understood, it was also adopted for the early stages of iCAM. However, the manner in which the F_L factor is used in CIECAM02 and iCAM are quite different. The shape of this luminance function is illustrated in Figure 135.

$$F_L = 0.2 \left(\frac{1}{(5L_A + 1)} \right)^4 (5L_A) + 0.1 \left(1 - \left(\frac{1}{(5L_A + 1)} \right)^4 \right)^2 (5L_A)^{\frac{1}{3}} \quad (45)$$

Figure 135. Shape of CIECAM02 F_L Function

The low-passed luminance image is used as input to Equation 45 to calculate this spatially varying exponent. The exponent is used to modify the IPT exponent function. This spatially varying exponent is very similar in nature to the low-pass masking function of the local contrast detection module described in Section 7.4 above, though it uses a single curve.

The IPT space serves as a uniform opponent color space, where I is the Lightness channel, P is roughly analogous to a red-green channel, and T a blue-yellow channel. Through a rectangular-to-cylindrical conversion it is possible to calculate chroma, and hue angle correlates. Equations 46-47 show this conversion.

$$C = \sqrt{P^2 + T^2} \quad (46)$$

$$h = \tan^{-1} \left(\frac{P}{T} \right) \quad (47)$$

The results of the iCAM calculation are pixel-by-pixel correlate of lightness, chroma, and hue. These color appearance correlates take into account the surrounding neighborhood, both in chromatic adaptation and local surround effects. Because of the spatial properties of iCAM, it is capable of predicting complex appearance phenomena such as simultaneous contrast, crispening, and spreading.

Simultaneous contrast (sometimes referred to as induction) causes the perception of a stimulus to shift in appearance away from the color of the background. This shifting away is described in terms of the opponent dimensions. Figure 136 illustrates a stimulus that exhibits simultaneous contrast in lightness (the gray square is physically identical on all three backgrounds) and its prediction by iCAM as represented by the iCAM lightness predictor. The prediction correctly increases the lightness of the gray square on the black background, while decreases the lightness of the gray square on the white background. This prediction is facilitated by the local adaptation features and local surround features of iCAM.

Figure 136. Prediction of Simultaneous Contrast Perception from iCAM. Original (a) and Appearance Prediction (b)

Crispening is the phenomenon whereby the color differences between two stimuli are perceptually larger when viewed on a background that is similar to the stimuli. Figure 137 illustrates a stimulus that exhibits chroma crispening and its prediction by the iCAM chroma correlate. The crispening prediction is also facilitated by the local chromatic adaptation features of iCAM.

Figure 137. Crispening Prediction of iCAM. Original (a) and iCAM Chroma Prediction (b)

Spreading is a spatial color appearance phenomenon in which the apparent hue of spatially complex image areas appears to fill various spatially coherent regions. Figure 138 provides an example of spreading in which the red hue of the annular region spreads significantly from the lines to the full annulus. The iCAM prediction of spreading is illustrated through reproduction of the hue correlate. The prediction of spreading in iCAM is facilitated by spatial filtering of the stimulus image, described in more detail in the following section.

Figure 138. Spreading Prediction of iCAM. Original (a) and iCAM Hue Prediction (b)

11.1 iCAM Image Difference Calculations

As iCAM evolved partially from the research leading up to the modular image difference framework, it can also serve to calculate metrics of image differences. The modules described above can easily be used in the iCAM framework. It is unnecessary to use the local contrast metric, since it is embedded into the model already. Figure 139 shows where the placement of the image difference modules inside the iCAM framework.

Figure 139. Using iCAM with Modules of Image Difference

The workflow for image difference calculations is similar to that for the standard iCAM model, using two images as input instead of a single. These images are processed through the spatially dependent chromatic adaptation transform before transformed into the IPT space. It is after the chromatic adaptation that the image difference modules are applied. Recall from Section 7.1 that the spatial filtering needs to be performed in an opponent color space. The IPT space itself can be used for these purposes, though it is necessary to not use the nonlinear IPT exponents (the 0.43 modulated with the CIECAM02 F_L) prior to performing the linear filtering. The spatial filtering, spatial frequency adaptation, and spatial localization are all performed in this linearized IPT opponent space, rather than the AC_1C_2 . These data are then transformed back to LMS signals for the IPT exponent contrast tone-reproduction. The images are finally converted back into the non-linear IPT space for color difference calculations.

Since the IPT space was designed to be a uniform color space, specifically in the hue dimensions, color differences can be calculated using a simple Euclidean distance metric as shown in Equation 48. Similarly, the color difference can be calculated from the cylindrical chroma and hue coordinates.

$$\Delta im = \sqrt{\Delta I^2 + \Delta P^2 + \Delta T^2} \quad (48)$$

The iCAM image difference predictions calculated using for the Sharpness Experiment and Print Experiment are shown in Figure 140 and 141 respectively.

Figure 140. iCAM Image Difference Predictions of Sharpness Experiment

Figure 141. iCAM Image Difference Predictions of Print Experiment

The iCAM image differences show correlations with the experimental data of 0.81 and 0.84 for the Sharpness and Print datasets respectively. It should not be surprising that these correlations are slightly worse than the cascaded image difference metric described above. The CIE color difference equations have had years of testing and refinement, while the IPT space has not been tested as a color difference formula before. It should be easy to extend the IPT color difference equations in similar manner as the CIE ΔE_{94} , if the IPT space turns out to be non-uniform in nature.⁷⁶ Another important consideration is the spatial filters, and specifically the opponent color space these filters are applied in. The filters described in

Sections 4.1 and 7.1 are based on the S-CIELAB equations, and are designed for use in a specific experimentally defined color space.³⁵ It might be necessary to slightly modify the filters for application in the IPT space.

The analysis of iCAM predictions can also be done taking into account the observer variability, as performed in Section 9.1.10. This analysis allows us to calculate the number of correct predictions, as well as the number of false negatives and positives. This can provide insight into the viability of using iCAM as a replacement, or augmentation into psychophysical experiments.

Figure 142. Experimental Predictions of iCAM with Observer Variability

Taking into account the observer variability iCAM predicted 10 false positives and 7 false negatives for sharpness data, and 3 false positives and 3 false negatives for the print experiment. This prediction is not as good as the best of the image difference models calculated using CIELAB and traditional color difference equations, but is close.

11.2 iCAM Summary

This section has introduced a first generation image color appearance model, iCAM, which is essentially a marriage of traditional color appearance models with the modular image difference framework described in this dissertation. This type of image appearance model forms a new direction for research, with the ultimate goal of a model capable of predicting spatially complex appearance correlates, such as lightness, chroma, hue, sharpness, graininess, and contrast. It is thought that these correlates can be used as a basis for a device-independent metric for image quality.

12 Conclusions and Future Directions

The fundamental focus of this dissertation can be described as the measurement of images, or specifically the measurement of the perception of images. One of the goals is the measurement of overall image quality. Image quality is ultimately a human response to, or perception of, spatially complex stimuli. Thus, the measurement of image quality is essentially a measurement of image appearance. This research has focused on the creation of computational models capable of predicting the perception of images, for use in image quality modeling.

12.1 Summary of Dissertation

Sections 1-5 outline some of the historical methods for measuring image perception and quality. These methods can be divided into two distinct categories, device-dependent and device-independent. Device-dependent image quality models correlate human perceptions such as sharpness, graininess, and contrast with specific imaging system attributes. This is accomplished through extensive psychophysical experimentation along with in-depth knowledge of the imaging system. These techniques have been used with varying degrees of success and are generally most useful when designing and evaluating complete systems. The device-dependent models, as the name implies, are only valid for a single imaging device. Change the device and a new image quality model must be developed.

Device-independent techniques use the information contained within the images themselves to characterize the imaging system so there is typically no need for knowledge of where the images came from for these types of models. These techniques generally rely on modeling of the human visual system to aid in prediction. The first types of device-independent models are threshold models of image differences, as described in Sections 3-5. These models are used to predict whether or not there is a perceived difference between two image pairs. This provides a first step towards modeling image quality because if an observer cannot see a difference between two images then the images must be of identical quality.

The next stage in device-independent modeling is the creation of a model capable of predicting magnitudes, and not just thresholds, of differences. It is towards this goal that the majority of this research has been focused. A modular framework for developing a color image difference model has been developed and described. This framework has been designed to be simple and flexible, and built upon traditional color difference equations. These traditional equations were designed to predict magnitude color difference for simple color patches on uniform backgrounds. The modular framework is designed to extend these models for spatial complex stimuli such as images. Several independent modules have been

described that account for spatial filtering, adaptation, and localization, as well as local and global contrast changes.

This framework has been tested using a series of psychophysical experiments that have been described in Section 8. A soft-copy sharpness experiment testing the effects of resolution, contrast, noise and spatial filtering was used to test the strengths and weakness of the individual modules. Two independent datasets were used to verify the sharpness experiment. The image difference metric was shown to be successful in predicting the general trend of these datasets.

The image difference framework is capable of predicting magnitudes of perceived differences but not directions of them. That is to say, the model is incapable of determining the cause of the difference, or whether or not the difference results in a better or worse image. To do this it is necessary to measure the appearance of the images. The independent nature of the modules in the image difference framework begins to allow for this type of measurement. The output from several of the individual the modules were shown to aid in the parsing of the causes of the perceived differences, such as changes in resolution, contrast, and spatial filtering.

The next stage in measurement is the ability to predict the appearance of, not just the difference between, images. An image appearance model extends traditional color appearance models in a similar manner that an image difference model extends traditional color difference equations. A first generation of an image appearance model, iCAM, was introduced in Section 11. iCAM evolved from the modular color image difference framework and the IPT uniform color space.

A model capable of predicting the appearance of spatially complex image stimuli can then be used in the final stage of image measurement, or the measurement of overall image quality. Just as a color appearance model adds the correlates of lightness, chroma, and hue to a color difference equation, an image appearance model adds correlates such as sharpness, contrast, and graininess. It is hoped that in the future this type of image appearance correlates can be used to create a device-independent metric for overall image quality.

12.2 Scientific Contributions

The scientific contributions of this study can be summarized as follows:

- Development of a framework for the creation of an image difference model. This framework was designed with three main goals: simplicity, modularity, and a basis on existing color difference research. The inspiration of this framework can be directly attributed to the S-CIELAB spatial extension to the CIELAB color difference equations.
- Development of specific modules for image difference framework. Several independent modules for the image difference framework were developed as a result of this research. These modules

include; spatial filtering based upon the human contrast sensitivity function, spatial-frequency adaptation that modifies the CSF for complex image stimuli, spatial localization to detect differences in edge information, and local and global contrast detection.

- Creation of a large bank of psychophysical experiment results. The sharpness experiment involved 140,000 observations on 288 images, by over 50 observers. All of the data generated by this experiment is freely available to the scientific community for any research use, and can be found at <http://www.cis.rit.edu/mcsl/>
- Development of a method for designing and evaluating image difference models. Techniques for qualitative and quantitative analysis of image difference model performance by direct comparison with psychophysical experiments were formulated. Typical research by the scientific community on the perception of image differences has not resulted in specific or rigorous model evaluations. By testing models on experimental results there can be a way to directly evaluate the performance of a model, or to directly compare models against each other. This can allow simultaneous model development by many researchers, with the ability to compare results.
- Development of a framework for an image appearance model. The combination of spatial properties of the human visual system with traditional color difference equations has lead directly to the concept of an image appearance model. A framework for such a model, iCAM, which combines spatial vision with color appearance has been developed.

It is important to stress that all of the data generated, and modules developed by this research are freely and easily available to the scientific community. The source-code for image difference modules, along with the source to the iCAM framework are available in several programming languages on the world-wide-web. This research was designed with the “open source science” mentality, to provides a set of tools for image difference and appearance research.

12.3 Conclusions and Future Directions

The primary focus of this research work was the formulation of the modular image difference framework. This framework, and the independent modules described, provides a new set of tools to the imaging community. The ability to predict the magnitudes of color differences for spatially complex images can be a valuable asset in the design of an imaging system. It could allow, for instance, a digital camera manufacturer to simulate the design of various components and determine whether the changes alter the perception of the resulting images, without the need to build the camera and perform extensive psychophysical research.

While totally replacing psychophysical research might be an ideal goal, in reality an image difference metric should be thought of as a tool for augmenting the human observer. Just as color difference equations have not replaced the need for a light-booth and visual examination, it is doubtful that an image difference metric would ultimately replace the need for visual examination.

The usefulness of any tool depends upon its proper application. A screwdriver makes a lousy hammer, as the saying goes. It is important to understand what the proper application of the image difference framework is. That can be accomplished through careful examination of its strengths and weaknesses. This research has focused on the relationship between the image difference metric predictions with actual psychophysical experiments. The ability of a metric to predict experimental results can provide insight into its usefulness as a tool.

This research has provided contributions to two distinct fields in the imaging community; the academic researchers working on modeling perceived image differences, and those interested in the practical application of image difference metrics. Previous research in image differences has focused mainly on qualitative model performance. Typical analysis includes looking at the histogram of color differences between pixel-by-pixel calculations and the image model calculations, or looking at a pair of images, and the resulting error image calculation. As long as the models made sense, e.g. the histogram was shifted towards lower color differences for half-tone images or the jpeg artifacts were visible in the error image, then the image difference model was doing its job. While these types of qualitative analysis are valuable in the development of the models themselves, they do not allow direct comparisons between previous and future models. By providing a dataset of psychophysical experiments to the community, as well as techniques for using the data, it is possible to directly compare models. This comparison can provide valuable insight into the relative strengths and weakness of each model, and can help accelerate future research.

Comparing image difference model predictions against real psychophysical data also can aid the practical application of these models, and provide guidelines towards their use. Correlation between the experimental data and the model predictions is important for direct model comparison, but it does not provide insight into how to apply these models in a real-world situation. No model can be expected to predict human observations better than the variation between the humans themselves. As such, it is possible to examine a model's ability to predict experimental results within these variations. This type of analysis provides insight into the number of "correct" predictions, as well as the number of false negatives or false positives. A correct prediction is a prediction that lies within actual human variation. A false negative is a prediction that is of a lower difference than a human would judge, while a false positive is a prediction that is greater than a human would judge.

The best combination of modules presented in this research was able to predict above 80% of the correct responses for most the three experiments described. Knowledge of this type of performance can guide the practical application of the image difference model, perhaps as a tool for augmenting actual human observations. For instance, a threshold image difference can be chosen to eliminate images that fall above from an experimental design. Depending on the application this threshold can be altered to minimize either false positives or false negatives. For instance, if the image difference model is used to choose various parameter levels for a new image compression algorithm the penalty for a false positive might be minimized. A false negative would mean that an image that is visibly too different than the non-compressed original would be included in the psychophysical experiment. This is not much of a problem, as that image would be detected in the experiment, though it does mean the experiment is larger than it needed to be. A false positive would mean that an image that is not visibly different than the original might be automatically eliminated, even though it is a valid image. By choosing an appropriate threshold level the false positives can be minimized. A small pilot experiment can be performed to obtain this threshold, or the threshold could be approximately 20% higher based upon previous research.

The image difference framework along with the methods for evaluating the framework, are designed to be tools for the imaging community. Like all tools, care must be applied before using them. This dissertation serves as a guide for using these tools.

12.3.1 Future Directions

Like all interesting research, this dissertation has provided insight into many questions and uncovered many more. The modular image difference framework provides a unique context upon which future research can be directly applied. One such example is the field of image appearance modeling. The iCAM image appearance model represents the first formulation of a color appearance model that takes into account spatial properties of the human visual system. The formulation of this model is a direct result of the lessons learned in the development of the modular image difference framework, and represents a focal point upon which future research can be concentrated. Current examples of this include extending the framework into the temporal domain for digital video applications⁸¹ and the rendering of high-dynamic range (HDR) images.⁸³

It is the hope that the frameworks presented as well as their methods for evaluation will represent a stepping-stone for rapid advancements towards the measurement of image perception. Making these tools available freely and easily to the scientific and industrial communities should only aid this advancement. All code and available data are available at <http://www.cis.rit.edu/mcsl>

13 References

1. P.G. Engeldrum, Image Quality Modeling: Where Are We?, *Proc of IS&T PICS Conference*, 251-255, (1999).
2. P.Engledrum, *Psychometric Scaling: A Toolkit for Imaging Systems Development*, Imcotek Press, Natick MA (2000).
3. M.D. Fairchild, Image Quality Measurement and Modeling for Digital Photography, *Proc. Of ICIS*, 318-319 (2002).
4. M.D. Fairchild, Measuring and Modeling Image Quality, *Chester F. Carlson Industrial Associates Meeting*, (1999).
5. B.W. Keelan, Characterization and Prediction of Image Quality, *Proc. of IS&T PICS Conference*, (2000).
6. B.W. Keelan, *Handbook of Image Quality: Characterization and Prediction*, Marcel Dekker, New York, NY (2002).
7. R.B. Wheeler, Use of System Image Quality Models to Improve Product Design, *Proc. Of IS&T PICS Conference*, (2000).
8. E.M. Granger and K.N. Cupery, An optical merit function (SQF), which correlates with subjective image judgments, *Photographic Science and Engineering*, **16**, 221-230 (1972).
9. P. Barten, Evaluation of subjective image quality with the square-root integral method, *Journal of the Optical Society of America A*, **7(10)**, 2024-2031 (1990).
10. P. Barten, *Contrast Sensitivity of the Human Eye and Its Effects on Image Quality*, SPIE Optical Engineering Press, Bellingham, WA (1999).
11. M. L. Giger, K. Doi, and C. E. Metz, Investigation of Basic Imaging Properties in Digital Radiography. 2. Noise Wiener Spectrum, *Medical Physics*, **11**, 797-805 (1984).
12. C. J. Bartleson, The combined influence of sharpness and graininess on the quality of colour prints, *Journal of Photo. Science*, **30**, 33- (1982).
13. E.M Granger, Specification of Color Image Quality, *Ph.D. Dissertation, University of Rochester*, (1974).
14. D.L. Lee, Performance of three image-quality metrics in ink-jet printing of plain papers, *Journal of Electronic Imaging*, **2**, 174-184 (1993).

15. S. Inoue, S. Yamazaki, N. Tsumura, and Y. Miyake, An Evaluation of Image Quality for Hardcopy Based on the MTF of Paper, *Journal of Imaging Science and Technology*, **44**, 188-195 (2000).
16. G.C. Higgins, Image quality criteria, *Journal of Applied Photographic Engineering*, **3**, 53-60 (1977).
17. Qun Sun, Spectral Imaging of Human Portraits and Image Quality, *PhD. Thesis, RIT* (2003).
18. S. Bouzit and L. MacDonald, Colour Difference Metrics and Image Sharpness, *Proc. IS&T/SID 8th Color Imaging Conference*, 262-267 (2000).
19. S. Daly, The Visible Differences Predictor: An algorithm for the assessment of image fidelity, Ch. 13 in *Digital Images and Human Vision*, A. B. Watson, Ed., MIT Press, Cambridge MA (1993).
20. J. Lubin, The Use of Psychophysical Data and Models in the Analysis of Display System Performance, Ch. 12. in *Digital Images and Human Vision*, A.B. Watson, Ed., MIT Press, Cambridge MA (1993).
21. Sarnoff Corp, JND: A Human Vision System Model for Objective Picture Quality Measurements, *Sarnoff Whitepaper*: <http://www.jndmetrix.com>, June (2001).
22. T. Ishihara, K. Ohishi, N. Tsumura, and Y. Miyake, Dependence of Directivity in Spatial Frequency Response of the Human Eye (2): Mathematically Modeling of Modulation Transfer Function, *OSA Japan*, **65** 128-133 (2002).
23. A.B. Watson, Visual Detection of spatial contrast patterns: Evaluation of five simple models, *Optics Express*, **6** 12-33 (2000).
24. A.B. Watson, The cortex transform: Rapid computation of simulated neural images, *Computer Vision Graphics and Image Processing*, **39**, 311-327 (1987).
25. F.W. Campbell and J.J. Kulikowski, Orientation selectivity of the human visual system, *Journal of Physiology*, **187**, 437-445 (1966).
 - A. Pantle and R.W. Sekuler, Contrast response of human visual mechanisms sensitive to orientation and direction of motion, *Vision Research*, **9**, 397-406 (1969).
26. J.A. Ferwerda, S.N Pattanaik, P. Shirley, and D.P. Greenberg, A Model of Visual Masking for Computer Graphics, *Proceedings of ACM-SIGGRAPH*, 249-258 (1996).
27. V. Volevich, K. Myszkowski, A. Khodulev and E. Kopylov, Using the Visual Differences Predictor to Improve Performance of Progressive Global Illumination Computations, *ACM Transactions on Graphics* **19**, 122-161 (2000).

28. B. Li, G. Meyer and R.V. Klassen, A Comparison of Two Image Quality Models, *SPIE Conf. on Human Vision and Electronic Imaging*, **3299**, 98-109 (1998).
29. E.D. Montag, and H. Kasahara, Multidimensional Analysis Reveals Importance of Color for Image Quality, *Proceedings of IS&T/SID 9th Color Image Conference*, 17-21 (2001).
30. P.J Burt and E.H. Adelson, The Laplacian pyramid as a compact image code, *IEEE Transactions on Communications*, **COM-31**, 532-540 (1983).
31. W.T. Freeman and E.H. Adelson, The design and use of steerable filters, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **13**, 891-906 (1991).
32. X.M. Zhang and B.A. Wandell, A spatial extension to CIELAB for digital color image reproduction, *Proceedings of the SID Symposiums*, **27**, 731-734 (1996).
33. E.W. Jin, X.F. Feng, and J. Newell, The Development of A Color Visual Difference Model (CVDM), *Proceedings of IS&T PICS Conference*, 154-158 (1998).
- A. B. Poirson and B. A. Wandell, The appearance of colored patterns: pattern-color separability, *J. Opt. Soc. A.*, (1993).
34. separability, *J. Opt. Soc. A.*, (1993).
35. X. Zhang, <http://white.stanford.edu/~brian/scielab/scielab.html>
36. X.M. Zhang, D.A. Silverstein, J.E. Farrell, and B.A. Wandell, Color Image Quality Metric S-CIELAB and Its Application on Halftone Texture Visibility, *IEEE COMPCON97 Digest of Papers*, 44-48 (1997).
37. X.M. Zhang and B.A. Wandell, Color image fidelity metrics evaluated using image distortion maps, *Signal Processing*, **70**, 201-214 (1998).
38. S.N Pattanaik, J.A. Ferwerda, M.D. Fairchild, and D.P. Greenberg, A Multiscale Model of Adaptation and Spatial Vision for Realistic Image Display, *Proceedings of ACM-SIGGRAPH*, 287-298 (1998).
39. S.N Pattanaik, M.D. Fairchild, J.A. Ferwerda, D.P Greenberg, Multiscale Model of Adaptation, Spatial Vision and Color Appearance, *Proceedings of IS&T/SID 6th Color Imaging Conference*, 2-7 (1998).
40. E.M. Granger, Uniform Color Space as a Function of Spatial Frequency, *Proceedings of the SPIE*, **1913**, 449-457 (1993).
41. M.D. Fairchild, *Color Appearance Models*, Addison Wesley, Reading MA, (1998).
42. S.L. Guth, Further applications of the ATD model for color vision, *Proceedings of the SPIE*, **Vol. 2414**, 12-26 (1995).
43. E.N. Dalal, R. Rasmussen, F. Nakaya, P.A. Crean, and M. Sato, Evaluating the Overall Image Quality of Hardcopy Output, *IS&T PICS Conference*, 169-173 (1998).

44. K. Miyata, N. Tsumuira, H. Haneishi, and Y. Miyake, Subjective Image Quality for Multi-Level Error Diffusion and Its Objective Evaluation Method, *Journal of Imaging Science and Technology*, **43**, 170-177 (1999).
45. CIE, Industrial Colour-Difference Evaluation, *CIE Publication 116-1995*, (1995).
46. CIE, Improvement to Industrial Colour-Difference Evaluation, *CIE Publication 142-2001*, (2001).
47. G.M. Johnson and M.D. Fairchild, Darwinism of Color Image Difference Metrics, *IS&T/SID 9th Color Imaging Conference*, Scottsdale, 108-112 (2001).
48. G.M. Johnson and M.D. Fairchild, A Top Down Description of S-CIELAB and CIEDE2000, *Color Res. Appl.* **27**, in press (2002).
49. K. Mullen, The contrast sensitivity of human color vision to red-green and blue-yellow chromatic gratings, *Journal of Physiology.*, **359** (1985).
50. G. J. C. Van der Horst and M. A. Bouman, Spatiotemporal chromaticity discrimination, *JOSA*, **59** (1969).
51. T. Movshon and L. Kiorpes, Analysis of the development of spatial sensitivity in monkey and human infants, *JOSA A*, **5** (1988).
52. E.D. Montag, *Personal Communication*, (2001).
53. B.A. Wandell, *Foundations of Vision*, Sinear Associates, Sunderland, MA (1995).
54. C.Blakemore and P.Sutton, Size adaptation: A new aftereffect, *Science*, **166**, 245-247 (1969).
55. M.A. Webster and E.Miyahara, Contrast adaptation and the spatial structure of natural images, *Journal of the Optical Society of America A*, **14** 2355-2366 (1997).
56. K. K. De Valois. Spatial frequency adaptation can enhance contrast sensitivity. *Vision Research*, **17**, 1057—1065 (1977).
57. J.S.Lee, Refined filtering of image noise using local statistics, *Computer Graphic and Image Processing* **15**, 380-389 (1981)
58. G. Westheimer, The spatial sense of the eye, *Inv. Ophthalmology Vision Science*, **18**, 893-912 (1979).
59. S.W. Kuffler, Discharge patterns and functional organization of mammalian retina, *Journal of Neurophysiology*, **16** , 37-68 (1953).
60. N. Maroney, Local Color Correction Using Non-Linear Masking, *Proc. of IS&T 8th Color Imaging Conference*, (2000).
61. R.C Gonzalez and R.F. Woods, *Digital Image Processing, 2nd Ed.*, (2001).

62. M. R. Luo, G. Cui, and B. Rigg, The development of the CIE 2000 Colour Difference Formula, *Color Research and Applications*, **26** (2000).
63. CIE, "The CIE 1997 Interim Colour Appearance Model (Simple Version), CIECAM97s," *CIE Pub. 131* (1998).
64. N. Moroney, M.D. Fairchild, R.W.G. Hunt, C.J Li, M.R. Luo, and T. Newman, The CIECAM02 color appearance model, *IS&T/SID 10th Color Imaging Conference*, Scottsdale, 23-27 (2002).
65. F. Ebner, and M.D. Fairchild, "Development and Testing of a Color Space (IPT) with Improved Hue Uniformity," *IS&T/SID 6th Color Imaging Conference*, Scottsdale, 8-13 (1998).
66. M.D. Fairchild, "A Revision of CIECAM97s for Practical Applications," *Color Res. Appl.* **26**, 418-427 (2001).
67. J.E. Farrell, Image quality evaluation, *Ch. 15 in Colour Imaging: Vision and Technology*, L.W. MacDonald and M.R. Luo, Eds., Wiley, Winchester, 285-314 (1999).
 - A. Vaysman and M.D. Fairchild, Degree of quantization and spatial addressability trade-offs in perceived quality of color images, *Color Imaging: Device Independent Color, Color Hardcopy, and Graphic Arts III, Proc. SPIE 3300*, 250-261 (1998).
68. J.Gibson, *Color Tolerances in pictorial images presented on various display devices*, RIT MS Thesis, (2002).
69. L.L Thurstone, A law of comparative judgment, *Psych Review*, **34**, 273-286 (1927).
70. P.Engledrum, *Psychometric Scaling: A Toolkit for Imaging Systems Development*, Imcotek Press, Natick MA (2000).
71. A.J. Calabria, *Compare and Contrast*, M.S. Thesis, RIT (2002).
72. C. Bartleson and Franc Grum, eds., *Optical Radiation Measurements, Vol 5: Visual Measurements*, Academic Press, Orlando Fl, (1984).
73. G.M. Johnson and M.D. Fairchild, On Contrast Sensitivity in an Image Difference Model, *Proc of IS&T PICS Conference*, 18-23, (2002).
74. R.S. Berns, *Billmeyer & Saltzman's Principles of Color Technology, 3rd Ed.*, John Wiley & Sons, New York, (2000).
75. NIST, http://www.itl.nist.gov/div898/strd/lls/lls_info.shtml, (2003).
76. Stata Software, <http://www.stata.com/support/faqs/stat/2sls.html>, (2003).
77. Allstat, <http://staff.bath.ac.uk/ccsphc/allstat.html>, (1996).
78. M.D. Fairchild and G.M. Johnson, Meet iCAM: a Next Generation Appearance Model, *Submitted to IS&T/SID 10th Color Imaging Conference*, (2002).

79. M.D. Fairchild and G.M. Johnson, Image Appearance Modeling, *Proc. Electronic Imaging, Santa Clara*, (2003).
80. N. Moroney, M.D. Fairchild, R.W.G. Hunt, C.J Li, M.R. Luo, and T. Newman, The CIECAM02 color appearance model, *IS&T/SID 10th Color Imaging Conference, Scottsdale*, 23-27 (2002).
81. G.M. Johnson and M.D. Fairchild, Rendering HDR Images, *Submitted to IS&T/SID 11th Color Imaging Conference*, (2003).

A. Psychophysical Results

The following Appendix shows the complete ranking of the paired comparison and rank order experiments described in the dissertation. The results shown are the non-normalized z-scores.

Sharpness Experiment: Combined Results

Image Name	Z-Score	Rank	Image Name	Z-score	Rank
300+10n+1.2+s	2.63	1	150+20n+s	0.13	37
300+20n+1.2+s	2.42	2	150+30n+s	0.08	38
300+1.2+s	2.21	3	150+30n+1.2+s	-0.04	39
300+1.1+s	2.05	4	150dpi	-0.04	40
300+10n+1.1+s	2.04	5	150+20n+1.2	-0.04	41
300+30n+1.2+s	2.01	6	150+20n+1.1	-0.04	42
300+10n+s	1.93	7	150+10n	-0.13	43
300+20n+1.1+s	1.86	8	150+30n+1.1+s	-0.35	44
300+1.2	1.83	9	150+20n	-0.47	45
300+10n+1.2	1.80	10	150+30n+1.2	-0.51	46
300+10n+1.1	1.77	11	150+30n+1.1	-0.57	47
300+1.1	1.74	12	150+30n	-0.90	48
300+20n+s	1.74	13	100+s	-1.11	49
300+20n+1.2	1.67	14	100+1.2+s	-1.29	50
300+30n+s	1.65	15	100+1.1+s	-1.42	51
300+20n+1.1	1.64	16	100+10n+1.2+s	-1.43	52
orig	1.57	17	100+10n+1.1+s	-1.47	53
300+30n+1.1+s	1.53	18	100+10n+s	-1.53	54
300+30n+1.1	1.39	19	100+20n+1.2+s	-1.68	55
300+20n	1.34	20	100+20n+1.1+s	-1.71	56
300+30n+1.2	1.34	21	100+1.1	-1.74	57
300+10n	1.32	22	100+1.2	-1.79	58
300+30n	1.08	23	100+10n+1.1	-1.81	59
150+10n+1.2+s	1.03	24	100+20n+s	-1.82	60
150+s	1.02	25	100+10n+1.2	-1.84	61
150+1.2+s	0.82	26	100+10n	-1.98	62
150+1.1+s	0.75	27	100dpi	-2.06	63
300+s	0.75	28	100+30n+s	-2.09	64
150+10n+1.1+s	0.72	29	100+20n+1.1	-2.10	65
150+20n+1.2+s	0.61	30	100+20n+1.2	-2.12	66
150+10n+s	0.46	31	100+20n	-2.19	67
150+1.2	0.42	32	100+30n+1.2+s	-2.23	68
150+1.1	0.41	33	100+30n+1.1+s	-2.32	69
150+10n+1.2	0.38	34	100+30n+1.2	-2.53	70
150+10n+1.1	0.33	35	100+30n	-2.66	71
150+20n+1.1+s	0.25	36	100+30n+1.1	-2.70	72

Sharpness Experiment: Cow Images

Image Name	Z-Score	Rank	Image Name	Z-score	Rank
300+1.2	2.50	1	150+10n+1.1+s	-0.04	37
300+20n+1.2+s	2.43	2	150+1.1+s	-0.04	38
300+10n+1.2	2.42	3	150+30n+1.2	-0.07	39
300+1.1	2.42	4	150+10n	-0.07	40
300+10n+1.2+s	2.37	5	150dpi	-0.14	41
300+10n+1.1	2.36	6	150+1.2+s	-0.15	42
300+20n+1.1	2.18	7	150+20n+1.1+s	-0.32	43
300+20n+1.2	1.92	8	150+20n	-0.37	44
300+30n+1.2	1.84	9	150+10n+s	-0.44	45
300+30n+1.1	1.76	10	150+30n	-0.52	46
300+30n+1.2+s	1.75	11	150+30n+1.1+s	-0.75	47
300+30n+s	1.58	12	150+20n+s	-0.78	48
orig	1.21	13	100+30n+s	-0.90	49
150+1.2	1.08	14	100+1.2+s	-0.92	50
150+s	1.06	15	100+1.1+s	-1.00	51
150+1.1	1.00	16	100+10n+s	-1.09	52
300+1.2+s	0.89	17	100+10n+1.2+s	-1.11	53
300+10n+1.1+s	0.89	18	100+10n+1.1+s	-1.21	54
300+1.1+s	0.88	19	100+1.1	-1.33	55
300+10n	0.86	20	100+20n+1.1+s	-1.37	56
300+10n+s	0.84	21	100+20n+s	-1.37	57
300+20n	0.81	22	100+10n+1.1	-1.40	58
150+10n+1.2+s	0.81	23	100+10n	-1.45	59
150+10n+1.1	0.80	24	100+1.2	-1.45	60
150+10n+1.2	0.74	25	100+10n+1.2	-1.55	61
150+30n+s	0.70	26	100dpi	-1.65	62
300+20n+1.1+s	0.69	27	100+20n+1.2+s	-1.66	63
300+20n+s	0.61	28	100+30n+1.2+s	-1.85	64
300+30n+1.1+s	0.59	29	100+30n+1.1+s	-1.87	65
150+20n+1.2	0.54	30	100+20n	-1.88	66
300+30n	0.47	31	100+20n+1.1	-1.88	67
150+20n+1.2+s	0.43	32	100+20n+1.2	-1.98	68
150+20n+1.1	0.42	33	100+30n	-2.08	69
100+s	0.01	34	300+s	-2.13	70
150+30n+1.1	-0.01	35	100+30n+1.2	-2.47	71
150+30n+1.2+s	-0.02	36	100+30n+1.1	-2.52	72

Sharpness Experiment: Bear Images

Image Name	Z-Score	Rank	Image Name	Z-score	Rank
300+10n+1.1+s	2.08	1	150+1.1	-0.09	37
300+20n+s	2.03	2	150+30n+1.1+s	-0.17	38
300+1.1+s	2.01	3	150+10n+1.2	-0.18	39
300+1.2+s	1.98	4	150+10n+1.1	-0.19	40
300+20n+1.1+s	1.97	5	150+1.2	-0.21	41
300+10n+1.2+s	1.92	6	150+20n	-0.25	42
300+10n+s	1.90	7	150+30n+1.2+s	-0.28	43
300+20n+1.2+s	1.89	8	150+20n+1.1	-0.53	44
300+s	1.66	9	150+20n+1.2	-0.54	45
300+30n+1.2+s	1.48	10	150+30n+1.1	-0.77	46
300+30n+1.1+s	1.38	11	100+1.2+s	-0.77	47
300+30n+s	1.31	12	100+10n+1.2+s	-0.82	48
orig	1.17	13	100+1.1+s	-0.82	49
150+10n+1.1+s	1.13	14	150+30n	-0.83	50
150+1.1+s	1.12	15	100+10n+1.1+s	-0.87	51
300+20n	1.11	16	100+s	-0.89	52
150+s	1.03	17	100+10n+s	-0.90	53
150+1.2+s	0.99	18	100+20n+1.2+s	-0.96	54
150+10n+s	0.97	19	100+20n+s	-0.96	55
300+10n	0.92	20	150+30n+1.2	-1.04	56
300+1.1	0.87	21	100+20n+1.1+s	-1.10	57
300+1.2	0.85	22	100+10n+1.1	-1.23	58
300+10n+1.1	0.84	23	100+10n	-1.33	59
150+20n+s	0.80	24	100+1.1	-1.41	60
300+30n	0.79	25	100+10n+1.2	-1.43	61
300+20n+1.1	0.78	26	100+1.2	-1.44	62
300+20n+1.2	0.70	27	100+20n+1.1	-1.44	63
150+10n+1.2+s	0.69	28	100+20n+1.2	-1.50	64
300+10n+1.2	0.63	29	100+20n	-1.55	65
150+20n+1.1+s	0.51	30	100dpi	-1.62	66
300+30n+1.1	0.51	31	100+30n+1.1+s	-1.71	67
150+20n+1.2+s	0.35	32	100+30n+s	-1.79	68
300+30n+1.2	0.19	33	100+30n+1.2+s	-1.85	69
150+30n+s	-0.03	34	100+30n	-2.23	70
150dpi	-0.06	35	100+30n+1.1	-2.30	71
150+10n	-0.08	36	100+30n+1.2	-2.37	72

Sharpness Experiment: Cypress Images

Image Name	Z-Score	Rank	Image Name	Z-score	Rank
300+1.2+s	2.31	1	150+10n+1.1	-0.07	37
300+10n+1.2+s	2.23	2	150+20n+1.2	-0.11	38
300+10n+1.1+s	2.14	3	150dpi	-0.19	39
300+10n+s	1.99	4	150+30n+1.2+s	-0.23	40
300+20n+1.1+s	1.95	5	150+10n	-0.26	41
300+20n+1.2+s	1.85	6	150+30n+1.1+s	-0.30	42
300+s	1.81	7	150+20n+1.1	-0.33	43
300+20n+s	1.78	8	150+20n	-0.41	44
300+1.1+s	1.74	9	150+30n+1.1	-0.47	45
300+30n+1.2+s	1.57	10	150+30n+1.2	-0.49	46
300+30n+1.1+s	1.52	11	150+30n+s	-0.56	47
300+1.2	1.43	12	150+30n	-0.78	48
orig	1.41	13	100+1.1+s	-0.81	49
300+10n+1.2	1.34	14	100+1.2+s	-0.89	50
300+30n+s	1.26	15	100+10n+1.2+s	-0.96	51
300+10n+1.1	1.22	16	100+10n+1.1+s	-1.06	52
300+20n	1.21	17	100+s	-1.07	53
300+20n+1.1	1.16	18	100+20n+1.2+s	-1.14	54
300+20n+1.2	1.15	19	100+1.1	-1.19	55
300+1.1	1.07	20	100+10n+s	-1.31	56
300+10n	1.05	21	100+10n+1.2	-1.35	57
150+1.2+s	0.99	22	100+20n+1.1+s	-1.37	58
300+30n+1.1	0.96	23	100+1.2	-1.40	59
150+1.1+s	0.91	24	100+10n+1.1	-1.53	60
300+30n+1.2	0.81	25	100+20n+1.2	-1.57	61
150+10n+1.2+s	0.76	26	100+20n+s	-1.62	62
300+30n	0.68	27	100+10n	-1.65	63
150+10n+1.1+s	0.66	28	100+20n+1.1	-1.75	64
150+s	0.56	29	100dpi	-1.76	65
150+20n+1.2+s	0.51	30	100+20n	-1.83	66
150+10n+s	0.45	31	100+30n+1.2+s	-1.86	67
150+20n+1.1+s	0.21	32	100+30n+s	-1.88	68
150+10n+1.2	0.18	33	100+30n+1.1+s	-1.91	69
150+20n+s	0.12	34	100+30n+1.2	-2.12	70
150+1.2	0.03	35	100+30n	-2.36	71
150+1.1	-0.04	36	100+30n+1.1	-2.43	72

Sharpness Experiment: Cypress Images

Image Name	Z-Score	Rank	Image Name	Z-score	Rank
300+1.2+s	2.28	1	150+20n+1.2	0.11	37
300+10n+1.2+s	2.22	2	150+10n+1.1	0.10	38
300+10n+1.1+s	2.17	3	150+10n	-0.10	39
300+20n+1.2+s	2.02	4	150+30n+1.2+s	-0.11	40
300+1.1+s	2.00	5	150dpi	-0.14	41
300+20n+1.1+s	1.85	6	150+30n+1.2	-0.17	42
300+10n+s	1.71	7	150+20n+1.1	-0.23	43
300+30n+1.2+s	1.68	8	150+30n+s	-0.29	44
300+s	1.66	9	150+30n+1.1+s	-0.30	45
300+20n+1.2	1.63	10	150+30n+1.1	-0.32	46
300+10n+1.2	1.58	11	150+20n	-0.42	47
orig	1.52	12	150+30n	-0.74	48
300+1.2	1.49	13	100+1.2	-1.06	49
300+30n+1.1+s	1.47	14	100+1.1+s	-1.11	50
300+10n+1.1	1.39	15	100+10n+1.2+s	-1.20	51
300+20n+s	1.34	16	100+10n+1.1+s	-1.22	52
300+1.1	1.25	17	100+1.2+s	-1.22	53
300+30n+1.1	1.14	18	100+20n+1.2	-1.32	54
300+20n+1.1	1.10	19	100+20n+1.2+s	-1.39	55
300+30n+1.2	1.08	20	100+20n+1.1	-1.40	56
300+30n+s	1.08	21	100+10n+s	-1.41	57
300+10n	0.96	22	100+10n+1.2	-1.43	58
300+20n	0.93	23	100+1.1	-1.45	59
150+1.2+s	0.81	24	100+10n+1.1	-1.50	60
150+10n+1.2+s	0.77	25	100+20n+1.1+s	-1.57	61
300+30n	0.56	26	100dpi	-1.60	62
150+1.1+s	0.53	27	100+s	-1.63	63
150+20n+1.2+s	0.38	28	100+10n	-1.73	64
150+10n+1.1+s	0.36	29	100+20n+s	-1.73	65
150+1.2	0.35	30	100+30n+1.2+s	-1.79	66
150+10n+1.2	0.33	31	100+20n	-1.86	67
150+20n+1.1+s	0.28	32	100+30n+1.2	-1.86	68
150+10n+s	0.28	33	100+30n+1.1+s	-1.90	69
150+1.1	0.20	34	100+30n+1.1	-2.15	70
150+20n+s	0.16	35	100+30n	-2.18	71
150+s	0.13	36	100+30n+s	-2.38	72

Contrast Experiment: Lightness Manipulation Z-Scores

Manip name	wakeboarder	veggies	pyramid	dinner	couple	Average
dec_sig_15	-0.77	-0.60	-0.90	-1.17	-1.34	-2.72
dec_sig_20	-0.21	-0.35	-0.53	-0.57	-0.80	-1.71
dec_sig_25	-0.22	-0.08	-0.35	-0.32	-0.20	-1.68
gma_0.900	-2.67	-2.61	-2.79	-3.08	-2.45	-1.22
gma_0.950	-1.74	-1.77	-1.77	-1.64	-1.63	-0.95
gma_1.00	0.12	0.22	0.08	0.33	-0.01	-0.74
gma_1.05	1.15	1.33	1.41	1.47	1.40	-0.49
hist_equal	1.24	1.38	1.42	-0.07	1.68	-0.31
inc_sig_10	1.53	1.44	1.58	1.91	1.71	-0.23
inc_sig_15	0.85	0.82	0.92	1.29	1.09	0.15
inc_sig_20	0.60	0.48	0.56	0.83	0.87	0.46
inc_sig_25	0.40	0.51	0.35	0.68	0.52	0.49
lin_0.0500	-0.34	-0.50	-0.24	-0.17	-0.32	0.67
lin_-0.0500	0.51	0.35	0.49	0.61	0.31	0.88
lin_0.100	-0.76	-0.80	-0.65	-0.70	-0.78	0.99
lin_-0.100	0.82	0.90	0.91	1.12	0.65	1.12
lin_0.150	-1.17	-1.22	-1.15	-1.32	-1.21	1.13
lin_-0.150	1.09	1.18	1.21	1.27	0.85	1.19
lin_0.200	-1.61	-1.71	-1.67	-1.79	-1.61	1.35
lin_-0.200	1.21	1.04	1.12	1.32	1.25	1.63

Contrast Experiment: Lightness Manipulation Z-Scores

Manip name	wakeboarder	veggies	pyramid	dinner	couple	Average
0s	-1.52	-1.39	-1.27	-1.59	-1.59	-1.47
25s	-0.87	-1.10	-0.67	-1.38	-1.00	-1.00
50s	-0.63	-0.84	-0.46	-1.24	-0.33	-0.70
75s	-0.08	-0.17	-0.11	-0.03	0.11	-0.06
100s	0.03	0.22	0.08	0.28	0.24	0.17
150s	0.59	0.78	0.67	0.93	0.84	0.76
200s	1.27	1.10	0.77	1.48	0.83	1.09
250s	1.13	1.40	0.99	1.55	0.90	1.19

Contrast Experiment: Chroma Manipulation Z-Scores

Chroma Scale	wakeboarder	veggies	pyramid	dinner	couple	Average
0	-0.86	-0.89	-0.82	-1.09	-0.80	-1.47
0.2	-1.54	-1.47	-1.41	-1.33	-1.73	-1.00
0.4	-0.56	-1.13	-0.80	-0.77	-0.85	-0.70
0.6	-0.41	0.09	-0.38	0.41	-0.28	-0.06
0.8	0.72	0.77	0.69	0.70	0.97	0.17
1	1.26	1.48	1.37	0.67	1.16	0.76
1.2	1.40	1.15	1.36	1.40	1.54	1.09

Print Experiment: Image QUALITY, Portrait, RIT Data

Image Manipulation	Rank	Z-Score
iso320 freq1 diam noise2	1	-0.78
iso320 freq2 rect noise0	2	-0.67
iso320 freq1 rect noise2	3	-0.66
iso320 freq2 rect noise2	4	-0.66
iso320 freq2 diam noise0	5	-0.63
iso320 freq1 rect noise0	6	-0.59
iso320 freq2 diam noise2	7	-0.56
iso320	8	-0.48
iso320 freq1 diam noise0	9	-0.31
iso1600 freq2 rect noise2	10	0.38
iso1600 freq2 diam noise0	11	0.45
iso1600 freq2 rect noise0	12	0.48
iso1600	13	0.51
iso1600 freq1 rect noise2	14	0.54
iso1600 freq1 rect noise0	15	0.56
iso1600 freq1 diam noise0	16	0.64
iso1600 freq1 diam noise2	17	0.84
iso1600 freq2 diam noise2	18	0.95

Print Experiment: Image SHARPNESS, Portrait, RIT Data

Image Manipulation	Rank	Z-Score
iso320 freq2 diam noise2	1	-1.08
iso320 freq1 diam noise2	2	-0.85
iso320 freq1 rect noise0	3	-0.76
iso320 freq2 diam noise0	4	-0.55
iso320 freq2 rect noise0	5	-0.54
iso320 freq1 diam noise0	6	-0.43
iso320 freq1 rect noise2	7	-0.41
iso320 freq2 rect noise2	8	-0.40
iso320	9	-0.16
iso1600 freq1 diam noise2	10	0.20
iso1600 freq1 rect noise0	11	0.41
iso1600 freq1 rect noise2	12	0.49
iso1600 freq2 diam noise2	13	0.60
iso1600 freq2 rect noise0	14	0.61
iso1600	15	0.68
iso1600 freq2 rect noise2	16	0.71
iso1600 freq1 diam noise0	17	0.73
iso1600 freq2 diam noise0	18	0.76

Print Experiment: Image GRAININESS, Portrait, RIT Data

Image Manipulation	Rank	Z-Score
iso1600 freq2 diam noise2	1	-0.96
iso1600 freq1 diam noise2	2	-0.94
iso1600 freq1 rect noise2	3	-0.59
iso1600 freq2 rect noise2	4	-0.57
iso1600	5	-0.54
iso1600 freq1 diam noise0	6	-0.46
iso1600 freq2 rect noise0	7	-0.41
iso1600 freq1 rect noise0	8	-0.39
iso1600 freq2 diam noise0	9	-0.38
iso320 freq1 diam noise2	10	0.30
iso320 freq2 diam noise2	11	0.48
iso320 freq1 rect noise2	12	0.58
iso320	13	0.60
iso320 freq2 rect noise2	14	0.61
iso320 freq2 rect noise0	15	0.62
iso320 freq1 diam noise0	16	0.62
iso320 freq1 rect noise0	17	0.65
iso320 freq2 diam noise0	18	0.77

Print Experiment: Image QUALITY, Portrait, Fuji Data

Image Manipulation	Rank	Z-Score
iso320 freq2 diam noise2	1	0.00
iso320 freq2 rect noise2	2	0.07
iso320 freq1 diam noise2	3	0.30
iso320 freq1 rect noise2	4	0.67
iso320	5	0.82
iso320 freq2 rect noise0	6	1.12
iso320 freq1 rect noise0	7	1.20
iso320 freq1 diam noise0	8	1.42
iso320 freq2 diam noise0	9	1.57
iso1600 freq1 diam noise0	10	3.46
iso1600	11	3.68
iso1600 freq1 rect noise0	12	3.90
iso1600 freq2 rect noise0	13	4.05
iso1600 freq2 diam noise0	14	4.43
iso1600 freq1 rect noise2	15	4.65
iso1600 freq2 rect noise2	16	4.88
iso1600 freq1 diam noise2	17	5.10
iso1600 freq2 diam noise2	18	5.17

Print Experiment: Image SHARPNESS, Portrait, Fuji Data

Image Manipulation	Rank	Z-Score
iso320 freq1 diam noise2	1	0.00
iso320 freq2 diam noise2	2	0.07
iso320 freq2 rect noise2	3	0.62
iso320 freq1 rect noise2	4	0.99
iso1600 freq1 diam noise2	5	1.14
iso1600 freq2 diam noise2	6	1.29
iso1600 freq2 rect noise2	7	1.83
iso1600 freq1 rect noise2	8	2.13
iso320	9	2.20
iso1600	10	2.66
iso320 freq1 rect noise0	11	2.66
iso1600 freq1 rect noise0	12	2.74
iso1600 freq1 diam noise0	13	2.81
iso320 freq1 diam noise0	14	2.88
iso1600 freq2 rect noise0	15	2.96
iso320 freq2 rect noise0	16	3.11
iso320 freq2 diam noise0	17	4.29
iso1600 freq2 diam noise0	18	4.37

Print Experiment: Image GRAININESS, Portrait, Fuji Data

Image Manipulation	Rank	Z-Score
iso320 freq2 rect noise0	1	0.00
iso320 freq1 diam noise0	2	0.00
iso320 freq2 diam noise0	3	0.07
iso320 freq1 rect noise0	4	0.30
iso320	5	0.67
iso320 freq1 rect noise2	6	1.30
iso320 freq2 rect noise2	7	1.68
iso320 freq2 diam noise2	8	2.22
iso320 freq1 diam noise2	9	2.45
iso1600	10	4.35
iso1600 freq2 rect noise0	11	4.42
iso1600 freq1 rect noise0	12	4.57
iso1600 freq2 diam noise0	13	4.57
iso1600 freq1 diam noise0	14	4.57
iso1600 freq1 rect noise2	15	4.87
iso1600 freq2 rect noise2	16	5.17
iso1600 freq1 diam noise2	17	6.52
iso1600 freq2 diam noise2	18	6.52

Print Experiment: Image QUALITY, Ship, RIT Data

Image Manipulation	Rank	Z-Score
iso320 freq1 rect noise2	1	-0.87
iso320 freq2 diam noise2	2	-0.85
iso320 freq2 diam noise0	3	-0.82
iso320 freq1 diam noise2	4	-0.65
iso320 freq1 diam noise0	5	-0.55
iso320	6	-0.45
iso320 freq2 rect noise2	7	-0.44
iso320 freq2 rect noise0	8	-0.40
iso320 freq1 rect noise0	9	-0.40
iso1600 freq1 rect noise0	10	0.52
iso1600 freq2 rect noise2	11	0.54
iso1600	12	0.54
iso1600 freq2 rect noise0	13	0.58
iso1600 freq1 rect noise2	14	0.59
iso1600 freq1 diam noise0	15	0.64
iso1600 freq2 diam noise2	16	0.64
iso1600 freq2 diam noise0	17	0.68
iso1600 freq1 diam noise2	18	0.69

Print Experiment: Image SHARPNESS, Ship, RIT Data

Image Manipulation	Rank	Z-Score
iso320 freq1 diam noise2	1	-1.21
iso320 freq2 diam noise2	2	-1.13
iso320 freq1 rect noise2	3	-0.81
iso320 freq2 diam noise0	4	-0.49
iso1600 freq2 rect noise2	5	-0.47
iso1600 freq1 diam noise2	6	-0.33
iso1600 freq2 diam noise2	7	-0.32
iso320 freq1 rect noise0	8	0.06
iso320 freq2 rect noise0	9	0.20
iso320 freq2 rect noise2	10	0.25
iso1600 freq1 rect noise2	11	0.28
iso320	12	0.29
iso1600 freq2 rect noise0	13	0.33
iso320 freq1 diam noise0	14	0.41
iso1600 freq1 rect noise0	15	0.53
iso1600 freq1 diam noise0	16	0.63
iso1600	17	0.83
iso1600 freq2 diam noise0	18	0.94

Print Experiment: Image GRAININESS, Ship, RIT Data

Image Manipulation	Rank	Z-Score
iso1600 freq2 diam noise2	1	-1.01
iso1600 freq1 diam noise2	2	-0.77
iso1600 freq1 rect noise2	3	-0.66
iso1600 freq2 rect noise0	4	-0.57
iso1600 freq2 rect noise2	5	-0.52
iso1600 freq1 diam noise0	6	-0.49
iso1600	7	-0.39
iso1600 freq2 diam noise0	8	-0.36
iso1600 freq1 rect noise0	9	-0.34
iso320 freq1 diam noise2	10	0.26
iso320 freq2 rect noise2	11	0.31
iso320 freq2 diam noise2	12	0.32
iso320	13	0.59
iso320 freq2 diam noise0	14	0.62
iso320 freq1 diam noise0	15	0.68
iso320 freq2 rect noise0	16	0.75
iso320 freq1 rect noise0	17	0.77
iso320 freq1 rect noise2	18	0.81

Print Experiment: Image QUALITY, Ship, Fuji Data

Image Manipulation	Rank	Z-Score
iso320 freq1 rect noise2	1	0.00
iso320 freq2 diam noise0	2	0.30
iso320 freq1 diam noise2	3	0.52
iso320 freq2 diam noise2	4	0.75
iso320 freq1 rect noise0	5	1.12
iso320 freq1 diam noise0	6	1.12
iso320	7	1.27
iso320 freq2 rect noise0	8	1.35
iso320 freq2 rect noise2	9	1.65
iso1600 freq2 rect noise2	10	2.19
iso1600 freq1 rect noise0	11	2.26
iso1600 freq1 rect noise2	12	2.33
iso1600 freq2 rect noise0	13	2.48
iso1600 freq2 diam noise0	14	2.48
iso1600	15	2.48
iso1600 freq1 diam noise0	16	2.63
iso1600 freq1 diam noise2	17	2.93
iso1600 freq2 diam noise2	18	3.23

Print Experiment: Image SHARPNESS, Ship, Fuji Data

Image Manipulation	Rank	Z-Score
iso320 freq1 diam noise2	1	0.00
iso320 freq2 diam noise2	2	0.22
iso320 freq1 rect noise2	3	0.68
iso320 freq2 diam noise0	4	0.76
iso1600 freq1 diam noise2	5	1.05
iso1600 freq2 diam noise2	6	1.43
iso1600 freq1 rect noise2	7	1.58
iso1600 freq2 rect noise2	8	1.65
iso320 freq1 rect noise0	9	2.58
iso320 freq1 diam noise0	10	3.12
iso320	11	3.20
iso320 freq2 rect noise0	12	3.27
iso320 freq2 rect noise2	13	3.65
iso1600 freq1 rect noise0	14	3.87
iso1600 freq2 diam noise0	15	4.10
iso1600 freq2 rect noise0	16	4.17
iso1600	17	4.32
iso1600 freq1 diam noise0	18	4.39

Print Experiment: Image GRAININESS, Ship, Fuji Data

Image Manipulation	Rank	Z-Score
iso320	1	0.00
iso320 freq2 rect noise0	2	0.15
iso320 freq1 diam noise0	3	0.37
iso320 freq2 rect noise2	4	0.37
iso320 freq1 rect noise0	5	0.52
iso320 freq1 rect noise2	6	1.06
iso320 freq2 diam noise0	7	1.21
iso320 freq1 diam noise2	8	2.14
iso320 freq2 diam noise2	9	2.29
iso1600	10	4.19
iso1600 freq1 diam noise0	11	4.34
iso1600 freq2 rect noise0	12	4.41
iso1600 freq2 diam noise0	13	4.56
iso1600 freq1 rect noise0	14	4.71
iso1600 freq1 rect noise2	15	5.64
iso1600 freq2 rect noise2	16	5.71
iso1600 freq2 diam noise2	17	6.53
iso1600 freq1 diam noise2	18	6.60

B. Pseudocode Algorithm Implementation

```
// A pseudocode representation of the modular image difference
// metric.

// First read in the RGB input images. Assume they are lossless
// Tiff images.
rgbImage1 = read_tiff('example1.tif')
rgbImage2 = read_tiff('example2.tif')

// Get the image size
imSizeX = size(rgbImage1, xDim)
imSizeY = size(rgbImage1, yDim)

// We must linearize the rgb images using a series of 3 1D luts
linRGBim1 = linearImage(rgbImage1)
linRGBim2 = linearImage(rgbImage2)

// Now convert the linearized RGB images into CIE 1931 XYZ using
// a 3x3 Matrix measured from a display

rgb2xyz = [[41.384, 22.155, .487], $
           [25.053, 51.424, 5.438], $
           [11.014, 9.743, 56.089]]

xyzImage1 = linRGBim1##rgb2xyz
xyzImage2 = linRGBim2##rgb2xyz

// The XYZ images will be used for the remainder of the analysis.

// We also need a transformation from XYZ tristimulus space to
// Wandell's AC1C2 space.

xyz2acc = [[ [278.7336, 721.8031,-106.5520], $
             [-448.7736, 289.8056, 77.1569], $
             [85.9513,-589.9859, 501.1089] ] / 1000.0
```



```

// Transform the XYZ images into ACC space

accImage1 = xyzImage1##xyz2acc
accImage2 = xyzImage2##xyz2acc

// Next we need to get the contrast sensitivity functions (CSF)
// Assume there is a function that returns the correct functions.
// We also need to know the cycles-per-degree of visual angle
// of the display device

cyclesPerDeg = 60

CSF = getCSF(imSizeX, imSizeY, cyclesPerDeg, /Movshon)

// There are several choices of contrast sensitivity functions
// such as Movshon, Daly, or Barten so the last flag would
// specify which function is desired

// If frequency boosting, aka Spatial Localization is desired then
// We can specify that now. Specify the location in CPD and width
// of a Gaussian boost

FreqBoost = getBoost(center=30, width=10)

// Cascade the CSF with the Freq Boost
CSF = CSF*FreqBoost

// Next we need to convert the ACC images into the frequency domain
// using a fast fourier transform

fftIm1 = fft(accImage1, /forward)
fftIm2 = fft(accImage2, /forward)

// The CSF can also be manipulated using spatial frequency adaptation
// at this point.

// for image independent adaptation we divide the luminance CSF by

```

```

(1/f)^(1/3)
CSF.luminance = CSF.luminance * CSF/(1/f)^(1/3)

// for image independent adaptation we first need to smooth the A channel of
// the image with a Lee filter, and raise that to an exponent
adapt1 = ( leeFilter(fftIm1.a) )^(1/3)
adapt2 = ( leeFilter(fftIm2.a) )^(1/3)

// Next we divide the luminance CSF by this adapt term
CSF1 = CSF2 = CSF
CSF1.luminance = CSF.luminance/adapt1
CSF2.luminance = CSF.luminance/adapt2

// and multiply the frequency image by the CSF

filtIm1 = fftIm1 * CSF1
filtIm2 = fftIm2 * CSF2

// Convert the filtered frequency image back to the spatial domain
filtACC1 = fft(filtIm1, /inverse)
filtACC2 = fft(filtIm2, /inverse)

// If we did not apply a frequency boost, we can perform the
// spatial localization in the ACC space using a high-pass filter
// such as the sobel

filtACC1 = sobel(filtACC1)
filtACC2 = sobel(filtACC2)

// This is also an ideal stage to perform the local contrast module
// if desired. This local contrast term uses a blurred version of the
// "A" channel to create a local series of tone reproduction curves
// based upon both localized and global contrast differences.

filtACC1 = localContrast(filtACC1)
filtACC2 = localContrast(filtACC2)

```

```

// The images need to be transformed back into CIE XYZ tristimulus
// values using the inverse of the matrix described above

filtXYZ1 = filtACC1##inverse(xyz2acc)
filtXYZ2 = filtACC2##inverse(xyz2acc)

// To calculate color differences we need to go into CIELAB space,
// and as such need a "whitepoint"

rgbWhite = [1, 1, 1]

xyzWhite = rgbWhite##rgb2xyz

// The XYZ images are then converted into CIELAB coordinates

labImage1 = xyz2lab(filtXYZ1, xyzWhite)
labImage2 = xyz2lab(filtXYZ2, xyzWhite)

// From the two CIELAB images we can calculate color differences using
// the CIE color difference equations. This creates an "error image" where
// each pixel represents the perceived color difference at that point.

errorAB = cieDeltaEab(labImage1, labImage2)
error94 = cieDeltaE94(labImage1, labImage2)
error2K = cieDeltaE2K(labImage1, labImage2)

// Finally, error stats can be calculated
meanError = mean(errorAB)
medianError = median(errorAB)
momentError = moment(errorAB)
stdev = sqrt(momentError[2])

```