High Dynamic Range (HDR) Display Perception

Fu Jiang
fj4136@rit.edu

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

Recommended Citation

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.
High Dynamic Range (HDR) Display Perception

by

Fu Jiang

A dissertation submitted in partial fulfillment of the
requirements for the degree of Doctor of Philosophy
in the Munsell Color Science Laboratory
College of Science
Rochester Institute of Technology

Mar, 2021

Signature of the Author

Accepted by

Coordinator, Ph.D. Degree Program
The Ph.D. Degree Dissertation of Fu Jiang has been examined and approved by the dissertation committee as satisfactory for the dissertation required for the Ph.D. degree in Color Science.

Dr. Mark D. Fairchild, Dissertation Advisor

Dr. Carl Salvaggio, External Chair

Dr. Michael J. Murdoch

Dr. Susan P. Farnand

Date
High Dynamic Range (HDR) Display Perception

by

Fu Jiang

Submitted to the
Munsell Color Science Laboratory
in partial fulfillment of the requirements
for the Doctor of Philosophy Degree
at the Rochester Institute of Technology

Abstract

Displays have undergone a huge development in the last several decades. From cathode-ray tube (CRT), liquid crystal display (LCD), to organic light-emitting diode (OLED), even Q-OLED, the new configurations of the display bring more and more functions into industry and daily life. In the recent several years, high dynamic range (HDR) displays become popular. HDR displays usually refer to that the black level of the display is darker and the peak being brighter compared with the standard dynamic range (SDR) display. Traditionally, the peak luminance level can be used as the "white" in characterization and calibration. However, for HDR displays, the peak luminance is higher than the traditional diffuse white level. Exploration of the perceptual diffuse white in HDR image when presented in displays is proposed, which can be beneficial to the characterizing and the optimizing the usage of the HDR display. Moreover, in addition to the “diffuse white”, 3D color gamut volume can be calculated in some specific color appearance models. Calculation and modeling of the 3D color gamut volume can be very useful for display design and better
characterizing display color reproduction capability. Furthermore, the perceptual
color gamut volume can be measured through psychophysical experiments. Com-
parison between the perceptual color gamut volume and the theoretical 3D gamut
volume calculations will reveal some insights for optimizing the usage of HDR dis-
plays. Another advantage of the HDR display is its darker black compared with the
SDR display. Compared with the real black object, what level of black is ‘perfect’
够 enough in displays? Experiments were proposed and conducted to evaluate that if
the HDR display is capable of showing “perfect” black for different types of back-
ground images/patterns. A glare-based model was proposed to predict the visual
“perfect” black. Additionally, the dynamic range of human vision system is very
large. However, the simultaneous dynamic range of human vision system is much
smaller and is important for the fine tuning usage of HDR displays. The simultane-
ous dynamic range was measured directly for different stimulus sizes. Also, it was
found that the simultaneous dynamic range was peak luminance level dependent.
A mathematical model was proposed based on the experimental data to predict the
simultaneous dynamic range. Also the spatial frequency effect of the target pattern
on the simultaneous dynamic range was measured and modeled. The four differ-
ent assessments about HDR displays perception would provide experimental data
and models for a better understanding of HDR perception and tuning of the HDR
display.
Acknowledgements

Firstly, I would like to express my sincere gratitude to my advisor Professor Mark D. Fairchild for the continuous support of my study and related research, for his patience, motivation, and immense knowledge. His guidance helped me in all the time of research and writing of this thesis. I could not have imagined having a better advisor and mentor for my study. Also, I am grateful for some of the wine knowledge from Dr. Fairchild.

Besides my advisor, I would like to thank the rest of my thesis committee: Dr. Carl Salvaggio, Dr. Susan Farnand, and Dr. Michael Murdoch. I appreciate Dr. Salvaggio being the committee chair in steering the committee and provide some suggestions from the other perspective. Additionally, I am very grateful to collaborating with Dr. Farnand and Dr. Murdoch in several projects during my PhD program. It is great honor to work with the two responsible and creative researchers. They have been very helpful in designing experiments, data analysis, and documentations.

Moreover, I would thank our lab mom, Val Hemink. She has taken care of all of our paperwork and being very generous to us during the whole PhD life. Moreover, I would like to express my gratitude to my colleagues, Anku, Lili Zhang, Matthew Ronnenberg, Hao Xie, Yongming Park, Luke Hellwig, Che Shen, Olivia Kuzio, Tucker Downs, Yue Yuan. Also, I would like to thank all observers participating my experiments. Additionally, my appreciation goes to my colleague and friends from my MS program, Zhaoyu Cui, Chi Zhang, Zichao Han, etc. It was a great time to have them at the beginning of my PhD program.

Finally, I would like to thank my family: my parents and my sister for supporting
me spiritually throughout writing this thesis and my life in general. It is not easy
to study alone overseas for years. Communications with the family and the support
from the family motivate me especially during the difficult times.
## Contents

1 Introduction ......................................................... 1  
   1.1 Organization .................................................... 3  

2 Background ......................................................... 5  
   2.1 Human Color Vision System ...................................... 6  
      2.1.1 Eye ........................................................... 6  
      2.1.2 Retina ....................................................... 8  
      2.1.3 Color Vision Theories ....................................... 9  
   2.2 Colorimetry ..................................................... 11  
      2.2.1 Tristimulus Values & Color Matching Function .......... 11  
      2.2.2 CIE Illuminants .............................................. 13  
      2.2.3 Chromaticity Diagrams ...................................... 14  
      2.2.4 Color Spaces ................................................ 14  
      2.2.5 Color Difference ............................................ 16  
      2.2.6 Metamerism .................................................. 17  

vii
2.3 Color Appearance Model ........................................ 19
2.4 Psychophysics ....................................................... 23

3 High Dynamic Range Display and Related Work .............. 25
  3.1 High Dynamic Display ........................................... 26
  3.2 HDR Color Appearance Models ................................. 33
    3.2.1 hdrCIELAB & hdrIPT ...................................... 33
    3.2.2 PQ & ICtCp ............................................. 37
    3.2.3 $J_z a_z b_z$ ............................................... 42
  3.3 Summary .......................................................... 44

4 Color Gamut Volume ................................................. 46
  4.1 Introduction ...................................................... 46
  4.2 Simulations ....................................................... 52
    4.2.1 Simulation Parameters ................................. 52
    4.2.2 Calculation Method ..................................... 53
    4.2.3 Simulation Result & Analysis ......................... 58
    4.2.4 Mathematical Model & Verification ................. 63
  4.3 Experiments ..................................................... 68
    4.3.1 Experiment I Design .................................... 68
    4.3.2 Experiment II Design .................................. 70
    4.3.3 Experiment Result Analysis .......................... 70
    4.3.4 Experiment Discussion ................................ 76
6.3.2 Glare Analysis .............................................. 111
6.4 Summary .................................................. 114

7 Direct Assessment of Human Visual System Simultaneous Dynamic Range
7.1 Introduction ................................................. 115
7.2 Experiment Setup .......................................... 118
  7.2.1 Experiment Design .................................... 118
  7.2.2 Apparatus .............................................. 119
  7.2.3 Stimuli ................................................... 120
7.3 Experiments & Results ..................................... 122
  7.3.1 Experiment I ........................................... 122
  7.3.2 Experiment II .......................................... 128
7.4 Modeling .................................................... 130
  7.4.1 Experiment I Modeling ............................... 131
  7.4.2 Experiment II Modeling ............................. 136
7.5 Discussion .................................................. 139
  7.5.1 Experiment I ........................................... 139
  7.5.2 Experiment II .......................................... 142
7.6 Conclusion ................................................ 145

8 Summary ..................................................... 147
  8.1 Future Work .............................................. 152
Appendices

A Slice Constant Lightness/Brightness

B Look Up Tables for Apple Pro Display

C Code to Optimize the Model for Simultaneous Dynamic Range

D Visual Search with Chroma Series on Different Background Colors

  D.1 Introduction
  D.2 Experiment Method
    D.2.1 Apparatus and Setup
    D.2.2 Task and Procedure
    D.2.3 Participants
  D.3 Result and Analysis
    D.3.1 Overall Performance
    D.3.2 Rectangle Layout Result
    D.3.3 Comparison between rectangle and “T”
    D.3.4 Searching Pattern Analysis
  D.4 Discussion
  D.5 Conclusions
  D.6 Acknowledgements
List of Figures

2.1 Anatomic diagram of the human eye [1]. ............................. 6
2.2 Organization of the retina [2]. ................................. 7
2.3 Distribution of rods and cons on retina [3]. ............................. 9
2.4 Illustration of color matching experiment. ....................... 12
2.5 Illustration of lighting metamerism. Different color masks over the lighting sources indicate different lights. .................. 17
2.6 Illustration of observer metamerism. Different color masks over observers/sensors indicate different observers/sensors. ............. 18
2.7 Example of simultaneous contrast. Two circles look the same under the same backgrounds (A). The two identical circles look different when under different backgrounds (B). ..................... 20
2.8 Example of crispening. The pair of patches on the three different backgrounds are the same. ............................. 21
3.1 Schematic digital light projector (up) and the real configuration (bottom). [17].

3.2 Peak luminance VS window size. Absolute luminance level is on the top, and normalized luminance is on the bottom [24].

3.3 Peak luminance heat map for background colors and window size [25].

3.4 Primaries of Rec. 709 (smallest triangle), DCI/P3 (mid-size triangle), and Rec. 2020 (largest triangle) in $xy$ chromaticity diagram (left) and in $u'v'$ chromaticity diagram.

3.5 Set up of experiment I (left) and experiment II (right) from Chen’s work [30].

3.6 CSF as a function of spatial frequency for different absolute luminance levels (A) and different visual sizes (B) [33].

3.7 Comparison of different encoding systems and with the Barten’s sensitivity model [34].

3.8 Example of testing example from [39].

3.9 Results of the color difference predictions [39].

4.1 Sampling points in $CIEXYZ$ in simulation of DCI/P3 with 500 cd/m$^2$ peak luminance.

4.2 Sampling points in JCh color space from CIECAM02 of DCI/P3 primaries with 200 cd/m$^2$ diffuse white setting.
4.3 Sampling points in JCh color space from CIECAM02 of DCI/P3 primaries with 200 cd/m^2 diffuse white setting, CLO/WLO = 0.5.

4.4 Sampling points in JCh color space from CIECAM02 of DCI/P3 primaries with 200 cd/m^2 diffuse white setting, CLO/WLO = 0.5.

4.5 Plots of 3D color gamut volume in CIELCh (left), JChDark (middle), QMhDark (right) against the peak luminance for the diffuse white as constant 200 cd/m^2. The solid dots are the simulated data and the dash line is the best fitting 3rd order polynomial function.

4.6 Color gamut volume in CIELCh (left), JChDark (middle), QMhDark (right) against the peak luminance for the diffuse white as 20% of the peak luminance. The solid dots are the simulated data and the dash line is the best fitting 3rd order polynomial function.

4.7 Normalized 3D color gamut volume in five combinations against CLO/WLO (different RGBW configurations). The solid lines are the best fitting as the 4th order polynomial functions.

4.8 Virtual primaries sets used for deriving the relationship between gamut volume against chromaticity area coverage.

4.9 Normalized color gamut volume against the chromaticy area coverage in xy (left) and u’v’ (right).

4.10 True color gamut volume against the predicted volumes for random 200 virtual primaries settings of random peak luminances with CLO/WLO as 0.1, 0.3, 0.5, 0.7, 0.9 and 1.
4.11 9 images for experiments, image 1 - 3 are natural images and image
4 - 9 are images with calibrated diffuse white level. .......................... 69
4.12 Perceptual color gamut volume against linear peak luminance (left)
and log peak luminance (right) for experiment I. ................................. 71
4.13 Perceptual color gamut volume (z-score) against log peak luminance
for experiment I. ................................................................................ 72
4.14 Perceptual color gamut volume against linear peak luminance (left)
and log peak luminance (right) for experiment II. ................................. 74
4.15 Perceptual color gamut volume (z-score) against log peak luminance
for experiment II. ................................................................................ 75
4.16 Normalized parameters against the perceptual color gamut volume
for experiment 1 (left) and experiment 2 (right). ................................. 78
4.17 Illustration of used volume by a pictorial image. The two black lines
present the two opponent channel in most color appearance models,
and the vertical axis (Z-axis) stands for the achromatic channel light-
ness/brightness. .................................................................................. 82

5.1 Configuration for experiment I and II, two black matte stands were
used to block any possible flare on the display. ................................. 88
5.2 6 images used for the three experiments. ................................. 89
5.3 Example of image#1 with five level gray patches on the center with
black padding. ..................................................................................... 90
5.4 Configuration for experiment III. ............................................. 91

5.5 Result of experiment I. The vertical axis is the absolute luminance in log scale. Gray, red, green, blue and magenta solid lines represent the five levels. Dash line represents the calibrated 200 nits diffuse white level. .................................................. 93

5.6 Comparison between 10 expert observers’ result (left) and 11 naive observers’ result (right) of experiment I. ................................. 94

5.7 Result of experiment II. Vertical axis is the log scale of normalized luminance level. Gray, red, green, blue and magenta solid lines represent the five gray patch levels. Dash line represents the calibrated diffuse white level, 20% of the peak luminance. .......................... 96

5.8 Comparison between 10 expert observers’ result (left) and 10 naive observers’ result (right) of experiment II. ................................. 97

5.9 Comparison between result of single-patch adjustment methodology and the five-level methodology for the image#1, #2, and #4. .... 98

5.10 Comparison of result of single-patch adjustment with three different patch sizes. size#1 is twice of that in five-level as in Figure 5.3, size#2 is four times of that, and size#3 is six times of that. .......... 100

5.11 Preview of original image#2 (top left), with size#1 patch (top right), with size#2 patch (bottom left), and size#3 patch (bottom right). 100

6.1 Original images used for the experiment. ................................. 105
6.2 The function of converting TSL sensor readout to the luminance Y (in log-log scale). The red line is the mean with green dash line indicating the 95% confidential interval. ................. 108
6.3 Configuration for measuring the $Y_{white}$, the calculating the $Y_{sample}$. 109
6.4 Example of the experimental image result. The horizontal axis is log10 scale of the luminance of the black circle and the vertical axis is the response, 0 or 1. ................................. 110
6.5 Summary of the glare, experimental result, luminance from the physical sample. ..................... 111
6.6 Glare and APL are plotted against the experimental threshold. Both axes are in log scale. The dash ellipse includes three obvious outliers for the glare against the threshold. ...................... 112
7.1 Illustration of experimental image. ......................... 118
7.2 Schematic top view of the experimental setup with the Apple XDR display on the black-cloth covered table along with two black foam boards. .............................. 121
7.3 Flowchart of the experiment. .............................. 124
7.4 Summary result of OBS1 with repeat observations. The 10% contrast threshold is plotted on the left and the simultaneous dynamic range is plotted on the right. Error bars represent the 95% confidential interval through bootstrapping. .............................. 125
7.5 Comparison between result of OBS1 and the average result of five observers. The measured threshold of 10% contrast is plotted on the left, and the simultaneous dynamic range is plotted on the right. Both horizontal axes are the stimulus size. Line types represents OBS1 and average observer. Line color represents different luminance levels of the bright stimulus. Error bars represent 95% confidence interval.

7.6 Result of experiment II for OBS1, the simultaneous dynamic range was plotted against the spatial frequency of the Gabor pattern for two bright stimulus luminance level settings.

7.7 Result of experiment I for single OBS1 with method 1 (log-linear) in solid lines, and method 2 (nonlinear) in dash lines.

7.8 Result of experiment I for single OBS1 with constrained fitting. Method 1 fitting (log-linear) is in solid lines, and method 2 (nonlinear) is in dash line.

7.9 Each data is plotted the mean with the 95% confidence interval. The solid lines are the best fittings for the four luminance levels. The dash lines are the optimized lines with constrain.

7.10 Best log parabola fitting for 452 cd/m² and 1600 cd/m² separately for varying spatial frequency patterns.

7.11 Best log-linear and nonlinear fittings for 452 cd/m² and 1600 cd/m² separately of the naive observers’ data.
7.12 Best log-linear fittings without constrain (solid lines) and with constrain (dash lines) for 452 cd/m² and 1600 cd/m² separately of the naive observers’ data. ........................................ 143

B.1 LUTs for different peak luminance levels of the Apple Pro Display. . . 179

D.1 Experiment setup. The bold black arrows are indicating the light directions for spot light and regular office lights. The walls are painted medium gray. ................................................. 188

D.2 CIEXYZ values of the measured flare over the 3 by 3 grid over the display. .................................................. 190

D.3 Rectangle layout (left) and “T” layout (right), the red objects are the potential target positions. ........................................... 191

D.4 Examples of test images for rectangle shape (left) and “T” shape (right). 192

D.5 Corresponding test images in the rotated set to the examples in Figure D.4. .......................................................... 192

D.6 Illustration of the three color scales with 1 CIEDE2000 between each consecutive patch. ....................................... 192

D.7 Schematic overview of experiment procedure. ................................ 196
D.8 Overall mean searching time for 2 second threshold (left) and 4 second threshold (right). The first row is the result of the normal vision observers and the second row is the result of the color vision deficient observers. The horizontal axis is color difference of the target from the mid-gray in unit of CIEDE2000, and the vertical axis is the searching time in second. Red lines are the fitting curves and the markers show mean and standard error.

D.9 Overall success rate for 2 second threshold (left) and 4 second threshold (right). The red solid line is the best-fitting for the normal vision observers’ data and the blue dash line is the best-fitting for the color vision deficient observers’ data.

D.10 Comparison of observers’ performance in the dark viewing environment (blue) and ambient viewing environment (red) based on mean searching time (left) and success rate (right) for 2 second threshold.

D.11 Comparison of success rate against color difference among the four backgrounds for the normal vision observers (left) and the color vision deficient observers (right) (rectangle shape only).

D.12 Success rate of different foreground chroma series for 2 second threshold of normal vision observers (left) and color vision deficient observers (right).

D.13 Success rate of rectangle and “T” layout for 2 second threshold (left) and 4 second threshold (right).
D.14 Gaze data of all observers at 0.1, 0.3, 0.5, 0.7 second from left to right
   of the rectangle (first row) and 'T' (second row). .................................. 211
D.15 Gaze data of random 3 observers colored in red, green and blue at
   0.1, 0.3, 0.5, 0.7 and 0.9 second of the “T” shape test images. ............... 212
D.16 Histogram of reaction time within 1 second for the dark and ambient
   viewing environments. ............................................................................ 214
D.17 Mean searching time of the three chroma series for 2 second threshold
   (left) and 4 second threshold (right). (color vision deficient observers,
   rectangle shape) ...................................................................................... 214
List of Tables

3.1 Constant parameters for $J_aJ_b J_c$. .................................. 44

4.1 Parameters for simulations. ............................................. 53

4.2 Values of $Rec2020_{constant}$ for the five combinations. ......... 64

4.3 Polynomial parameters for calculating $PeakLumRatio$. ............ 64

4.4 Polynomial parameters for calculating $xyRatio$. .................... 65

4.5 Polynomial parameters for calculating $RGBW Ratio$. ............... 66

4.6 Mean errors of the verification simulations. .......................... 67

4.7 $R^2$ between the parameters and the perceptual color gamut volume. . 78

6.1 $R^2$ between threshold, threshold+bk and different parameters. .... 113

7.1 Optimized parameters for log linear and nonlinear fitting over experiment I data. ........................................... 132

7.2 Optimized parameters under constrain for log linear and nonlinear fitting over experiment I data. ................................. 134
7.3 RMSE in log unit, the same for $L_{\text{threshold}}$ and the simultaneous dynamic range, of the two methods, for no constrain and constrained conditions. .................................................. 135

7.4 Parameters for the glare-based model: bright stimulus luminance level dependent $k$ and $b$ Eq. 7.6, and $k$ for Eq. 7.7. ......................... 136

7.5 Optimized parameters for log parabola fitting over experiment II data. 139

7.6 RMSE in log unit, the same for $L_{\text{threshold}}$ and the simultaneous dynamic range, of the two methods with constrain for average observer’s data. .................................................. 140

7.7 RMSE in log unit, the same for $L_{\text{threshold}}$ and the simultaneous dynamic range, of the fitting results for average observer’s data against the estimated glare (in Figure 7.12). ......................... 142

D.1 CIEXYZ of rectangle target for the dark and ambient viewing environments (Y is in units of cd/m$^2$). ................................. 194

D.2 CIEXYZ of “T” target for the dark and ambient viewing environments (Y is in units of cd/m$^2$). ........................................ 195

D.3 CIEXYZ of four backgrounds (Y is in units of cd/m$^2$). .................. 196

D.4 ANCOVA analysis result for rectangle shape layout. Bold data are those p values below the significant level, 0.05. ......................... 204

D.5 ANCOVA analysis result for “T” shape layout. Bold data are the p values lower than the significance level, 0.05. ......................... 210
D.6 Parameters for background effects (shaded for gray background, plain white for the average of the other three). . . . . . . . . . . . . . . . . 218

D.7 Probability of false rejection for two different layouts and the joint of two layouts). . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 220
Chapter 1

Introduction

The display has become an important tool in daily life in delivering images and videos. From the cathode-ray tube (CRT) display to the current LCD display, display becomes more efficient and larger. Recently, high dynamic range (HDR) displays become more popular.

Compared with the tradition standard dynamic range (SDR) display, HDR displays provide a higher dynamic range, “blacker” on the black level and brighter on the peak luminance. The SDR signal standard allows maximum 100 cd/m² peak luminance with a requirement of black level less than 0.5 cd/m². The current commercial HDR display can easily achieve a 600 cd/m² peak luminance and a black level ≤0.01 cd/m². Also along with the increasing dynamic range, the HDR display usually includes a wider color gamut presenting more vivid and colorful image/video. The traditional Rec. 709 adopts the sRGB color gamut primaries. Now, the OLED display can achieve DCI/P3 color gamut coverage easily. Also, quantum dot (QD) technology, which works like a spectral filter on the light emitting diode, could further expand the primaries coverage on the chromaticity diagram.

Color gamut area over the chromaticity diagram is usually used to illustrate the display’s ability in color reproduction. However, the real color gamut used to display colors is the 3D volume rather than the 2D chromaticity diagram. The 2D chromaticity diagram is more of an easy way in illustrating the hues of the primaries but not the real primaries. It is
necessary to explore the 3D color gamut volume. The first part of this thesis focuses on the 3D color gamut volumes expressed in color appearance spaces, including a novel calculation method, analysis, a mathematical model and an independent validation.

The 3D color gamut volume of a display is the computational volume in a designated 3D space. It is meaningful for manufacturers to better understand and explain the color reproduction ability of the display. However, it may not be the same in terms of the visual experience in the perspective of the normal observers or the consumers. Therefore, it is also important to evaluate the display’s color reproduction ability in the perspective of the normal observers. There is very limited related work so far. Therefore, in addition to the computational 3D color gamut volume, two color gamut volume evaluation experiments were conducted and analyzed for evaluating displays’ color reproduction ability in the perspective of normal observers.

Another important question for HDR display is the diffuse white level. Diffuse white is a fundamental concept in almost all color appearance models, where the diffuse white is used as an anchor point, and is important for display characterization. For SDR display, the peak luminance level is usually used as diffuse white in the characterization. When displaying images, the diffuse white is close to the peak luminance level of the SDR display. However, for the HDR display, the diffuse white level is not necessarily to be the peak luminance, and usually it is much lower than the peak luminance. The peak luminance would be too high to be considered as the diffuse white level as the high luminance range are mostly used for highlights, specular reflection, direct lighting, etc. It would be necessary to explore the diffuse white level for the practical usage of the HDR display. On the HDR capture side, the diffuse white would be important for the calibration and the exposure setting.

As introduced before, a “deeper” black level compared with the SDR display is another important feature of the HDR display. There have been many studies about how “black” is good enough. It varies a lot among the experiments’ designs and the applications. The result could also depend on the display. With the development of HDR displays, they can achieve
a much deeper black, especially the OLED HDR display. It is worth exploration about the black level of the HDR display, i.e. how black is enough, how that varies with patterns and image, and if those are predictable. This can be considered as visually ‘perfect’ black. Experiments were designed and conducted to measure the visual ‘perfect’ black level under different backgrounds and different patterns. A model was built to predict the ‘perfect’ black level.

Lastly, the simultaneous dynamic range was measured. The simultaneous dynamic range refers to the dynamic range when the bright and dark area are presented simultaneously to the observers. It is known that human visual system can adapt to a 14 log unit luminance level, where $10^{-6}$ cd/m$^2$ to 1 cd/m$^2$ is called scotopic vision and 0.01 cd/m$^2$ to $10^{8}$ cd/m$^2$ is called photopic vision. However, the simultaneous dynamic range will be much lower due to two reasons: 1) restriction on adaptation, the bright region would prevent the complete adaptation on the dark region, and 2) the glare caused by the bright region on the visual system will impair the discrimination capability in the dark region as well. There has been research on the simultaneous dynamic range, mainly through physiological measurement, i.e. the response of the photoreceptors. The recent psychophysical measurement was not a direct measurement. Therefore, this work from this study would provide more psychophysical measurement data and be beneficial to some HDR applications, especially some fine-tuning and mapping method, optimization of HDR displays.

1.1 Organization

In this thesis, the goal is to explore the HDR display in four different aspects: the 3D color gamut volume, the diffuse white level, the ‘perfect’ black level in HDR displays, and the simultaneous dynamic range. In the background, Chapter 2, some general concepts about the human vision perception, colorimetry, color appearance model, and psychophysics are introduced and explained. In chapter 3, HDR display technologies, and proposed color spaces and
models for HDR will be presented, explained and summarized. In Chapter 4, the 3D color gamut volume in the color appearance models will be calculated and modeled. Moreover, the perceptual color gamut volume will be explored through several experiments. In Chapter 5, the perceptual diffuse white in HDR images will be explored through several psychophysical experiments. The result would be helpful in HDR video/image capture and the color appearance calculations of images on the HDR display. In Chapter 6, the work exploring the black level on the display, mainly measuring the JND step from the ‘perfect’ physical black sample on different background images/patterns, is presented. A glare-based model is proposed to predict the “perfect” visual black. In Chapter 7, the simultaneous dynamic range work was presented. The simultaneous dynamic range was measured for different peak luminance levels and different stimulus sizes. A mathematical model, predicting the simultaneous dynamic range for different stimulus size, is proposed based on experimental data, and tested by some independent data. Finally, Chapter 8 will summarize the four different aspects, and future work, along with the publications from these studies. In the Appendix, there are some Matlab code used in the work and one independent study from the PhD program, “Visual Search with Chroma Series on Different Background Colors”, is included.
Chapter 2

Background

This dissertation is focusing on the HDR display perception in four different aspects. It includes a theoretical color gamut volume, the perceptual color gamut volume, the diffuse white level in HDR image, the “perfect” black level in displays, and the simultaneous dynamic range. The 3D color gamut volumes are the 3D volume in the color spaces. It is necessary to introduce some color spaces concepts in the background. Moreover, in addition to the color spaces, the color appearance models are developed to predict more visual effects. Also, the color gamut volume in color appearance model spaces is more meaningful. Additionally, the psychophysical experiments are a typical way in measuring the observers’ perceptual color gamut volume. Therefore, psychophysics is also included in the background. To sum up, the background chapter includes four different parts: 1) a general introduction of human color vision system (section 2.1); 2) basic concepts about colorimetry (section 2.2); 3) color appearance models (section 2.3); 4) psychophysics (section 2.4).
2.1 Human Color Vision System

2.1.1 Eye

Figure 2.1 shows the anatomic diagram of the human eye with labels of the key components. The cornea is the transparent part of the front of the eye. It is the first structure of human vision system that interacts with incident light. The iris is the structure controlling the pupil, which is similar as aperture of a camera. Eye color is usually defined by the iris. After the pupil, the light should go through the lens. The lens is a flexible structure with varying optical refraction index. The center of the lens has higher refraction index than the edge. The vitreous gel is the clear gel filling the space between lens and retina. It does no affect human's perception of color. The retina is another key structure in the eye having significant impact on human color perception. It is a thin layer of cells located at the back of the eye.
Figure 2.2: Organization of the retina [2].
2.1.2 Retina

Figure 2.2 plots the organization of the retina. The light goes through the front layers of cells. The rods and cones are the photoreceptors, which are excited by photons and emit electrons into the cell system. The pigment epithelium is a layer mainly serving to absorb photons passing through the rods and cones. This layer could reduce the scattering in the rods and cones. The electronic signals from the rods and cones transmit into bipolar cells directly. Usually each bipolar cell connects with multiple photoreceptors. One higher layer would be the ganglion cells, which take multi bipolar cells as input. In addition to this vertical structure, the horizontal cells and the amacrine cells are also important types of cells. Horizontal cells usually span across photoreceptors and summate the signal before passing to other photoreceptor cells. Amacrine cells have similar structure as horizontal cells but operate at the inner nuclear layer. Amacrine cells connect the bipolar cells and affect the synapses between bipolar cells and ganglion cells. Finally, the axons of the ganglion cells comprise the optical nerve fiber. The nerve fiber layer is collected through the optical nerve (Figure 2.1). The connections between these types of photoreceptor cells are much more complex.

Another interesting fact about the photoreceptors is their distribution. As mentioned above, there are two types of photoreceptors: rods and cones. Rods and cones mainly can be distinguished by their function that the rods serve at low luminance level and the cones serve at higher luminance level. Another interesting distinction between rods and cones is the spatial distribution. Figure 2.3 illustrates the rods’ and the cones’ distribution across the retina. Cones are usually concentrated on the fovea region and are much more sparsely distributed on the periphery. The fovea lacks rods. While the density of the rods is the highest on the side of fovea and gradually decreases onto the periphery. One thing that should be noted is that there is a blind spot around 16 degree from the fovea, where the optical nerve is located (see Figure 2.1).
2.1.3 Color Vision Theories

The well known color vision theory is trichromatic theory. This was developed based on the work of Maxwell, Young, and Helmholtz, therefore it is also called Young-Helmholtz theory. The theory assumed that three different images were formed based on the three types of receptors. The three images were then transmitted into the brain for further processing. Of course, there is no doubt about the three types of receptors. Another theory was proposed by Hering [4]. This theory was based on some subjective observations. An interesting finding was how color was described, that there was no reddish-green, yellowish-blue. Hering, therefore, proposed the opponent color theory, that the color was defined by red-green, yellow-blue, and light-dark. In modern days, the anatomy and some measurement methods verified and measured the spectral sensitivity of the three receptors and the opponent theory was modified more precisely as the signal difference between the three types of receptors. Also, the purified hue was not exactly opponent against each other.

In addition to the color vision theories, there is another important mechanism in human vision system: adaptation. The human vision system automatically adapt the sensitivity of
the photoreceptors under the viewing environment. This mechanism allows for the human ability to see a wide range of absolute luminance level, from sunny bright outdoor to dim night indoor. Besides the light-dark adaption, human vision system would also adapt to colors. For example, if human vision was exposed to a uniform color area for a period of time, vision system would see an opponent color of that when looking away at a grey or white area. This is an important mechanism in explaining many human color vision effects.

Human color vision also has some spatial and temporal characteristics. Spatially, any signal can be expressed as a combination of sinusoidal signals, which is also known as Fourier transformation. The human color vision system has a response to the pattern with different spatial frequency. This is close to Modulation Transfer Function (MTF) in optical systems, which has different modulation to different frequency signals. In vision system, it is usually explored through measuring the response of human vision system over the certain contrast with different frequencies. So, this is also called contrast sensitivity function (CSF). The CSF of the human vision system is close to a low pass filter with a peak response around 2-5 cycle/degree, and decreases gradually to zero at 60 cycle/degree [5]. This is the CSF of the luminance channel or the CSF to the black-white contrast pattern. The CSF to red-green pattern or yellow-blue can also be measured using the same method [6]. The human vision system’s CSF of the color channels demonstrates similarity and differences compared with the CSF of the luminance channel. First, CSFs of red-green and yellow-blue are also a low pass filter shape. However, CSFs of both color channels have their peaks at 0 cycle/degree, and decrease gradually. Both have much lower cutoff frequencies compared with that of the luminance channel. The temporal CSF shares some similarities with the spatial CSF. The temporal CSF of the luminance channel also has its peak sensitivity at 3-4 Hz, while the temporal CSFs of the chromatic channels have their peaks at lowest frequency. Also, the chromatic temporal CSF has a lower cutoff frequency than the luminance temporal CSF. Of course, the curves can shift with the shape and the size of the measuring stimulus [7].
2.2 Colorimetry

Color is a sensation essentially. For a precise color reproduction and exchange of colors, it is necessary to precisely describe color. Colorimetry is the measurement of colors. A short summary of some important concepts will be introduced.

2.2.1 Tristimulus Values & Color Matching Function

In colorimetry, a color is usually described and expressed as a set of three values mathematically. The direct understanding of the three values would be the three responses from the three types of cones. For example, the three types of cones will have three responses to a certain object. These three values can be called tristimulus values. Eq. 2.1 is the equation to compute the tristimulus values/cones responses for any stimulus. \( LMS \) are the cones responses. \( \lambda_1 \) and \( \lambda_2 \) are the spectral boundaries. Usually \( \lambda_1 \) and \( \lambda_2 \) are 360\( nm \) and 780\( nm \). \( \Phi_{\lambda_i} \) is the spectral intensity from a stimulus at a certain wavelength \( \lambda_i \). \( L_{\lambda_i} / M_{\lambda_i} / S_{\lambda_i} \) are the cone sensitivities at \( \lambda_i \). \( L_{\lambda_i} / M_{\lambda_i} / S_{\lambda_i} \) are called color matching functions since stimuli with the same \( LMS \) induce the same responses of human color vision system.

\[
L = \int_{\lambda_1}^{\lambda_2} \Phi_{\lambda_i} \ast L_{\lambda_i} \ast d\lambda_i \\
M = \int_{\lambda_1}^{\lambda_2} \Phi_{\lambda_i} \ast M_{\lambda_i} \ast d\lambda_i \\
S = \int_{\lambda_1}^{\lambda_2} \Phi_{\lambda_i} \ast S_{\lambda_i} \ast d\lambda_i
\] (2.1)

In addition to the cones sensitivities, there are many matching functions that meet the requirement that the computed matched tristimulus values from stimuli should be matched for average observers. Actually, for any three primaries, the color matching functions can be derived through a color matching experiment. Figure 2.4 illustrates the color matching experiment. In the figure, \( R-G-B \) are the primaries. \( \lambda_i \) means the monochromatic stimulus.
For any three primaries, observers can match the colors of the two semicircles. The matching results can be derived by Eq. 2.2. Therefore, a series of monochromatic stimuli can result in a spectral matching results, which are color matching functions for these three primaries. For example, CIE (International Commission on Illumination) proposed RGB color matching functions for three monochromatic primaries as 435.8\text{nm}, 546.1\text{nm}, and 700.00\text{nm}. The more well-known color matching functions are XYZ from CIE, which is a transform of \(R-G-B\) color matching functions. The main motivation of proposing \(XYZ\) is to eliminate the negative values from the \(RGB\) color matching functions. \(XYZ\) color matching functions, Eq. 2.3, have an additional scale \(k\) compared with the response function in Eq. 2.1. Here, \(k\) is used to normalize the \(Y\) to 100 for the lighting source shown in Eq. 2.4. For a reflective object, the reflectivity of an object does not change if the intensity of the lighting changes. This, \(k\), is used to prevent the effect of using lighting source with different intensities.

\[
\begin{align*}
R_L + G_L + B_L + \Phi_{\lambda_i} &= R_R + G_R + B_R \\
\Phi_{\lambda_i} &= (R_R - R_L) + (G_R - G_L) + (B_R - B_L)
\end{align*}
\]
\[ X = \int_{\lambda_1}^{\lambda_2} k \cdot \Phi_{\lambda_i} \cdot x_{\lambda_i} \cdot d\lambda_i \]
\[ Y = \int_{\lambda_1}^{\lambda_2} k \cdot \Phi_{\lambda_i} \cdot y_{\lambda_i} \cdot d\lambda_i \]  \hspace{1cm} (2.3)
\[ Z = \int_{\lambda_1}^{\lambda_2} k \cdot \Phi_{\lambda_i} \cdot z_{\lambda_i} \cdot d\lambda_i \]

\[ k = \frac{100}{\int_{\lambda_1}^{\lambda_2} S_{\lambda_i} \cdot y_{\lambda_i} \cdot d\lambda_i} \]  \hspace{1cm} (2.4)

### 2.2.2 CIE Illuminants

CIE has proposed a number of spectral power distribution as CIE illuminants for colorimetry. The standard illuminants include A, C, D65, D50, F2, F8, and F11. Before we introduce these standard illuminants, an important concept should be explained, color correlated temperature (CCT). This is an concept to describe a lighting source. The black body with a certain temperature has a specific power distribution, known as the Planck function. The power distribution will have the corresponding tristimulus values. The temperature of the blackbody having the closest \( XYZ \) to the lighting is called the CCT of the lighting source.

**Illuminant A** has a CCT of 2856K. This is usually used when the incandescent illumination is of interest. **CIE Illuminant C** is a modified by a filter of illuminant C providing a power distribution with CCT of 6774K. **CIE illuminant D65 and D50** are two examples from the CIE D-series of illuminants, which were built based on daylight measurements. These two are the most used in colorimetry as they are used as defined white points for many standards. **CIE illuminant F** is a series of power distributions representing fluorescent sources, with different CCTs. For example, **F8** has the CCT of 5000K but a different power distribution from D50. More details of these can be found in the CIE reports.
2.2.3 Chromaticity Diagrams

The color of a stimulus is specified by the tristimulus values. For convenience, two dimensional representations of colors were developed as chromaticity diagrams. A projection method is used to project the three dimensional tristimulus values into the 2D plane. There are two chromaticity diagrams developed by CIE, $xy$ and $u'v'$. CIE $xy$ is simply normalized XY by sum of XYZ, shown in Eq. 2.5. Chromaticity diagram has been used often to illustrate colors, color difference, and color gamut. It should be noted that even through the chromaticity diagram provides an easy way in showing colors, it discards one dimension of the colors. Therefore, any quantization property on the chromaticity diagrams, i.e. color difference, should be examined before being used. CIE established $u'v'$ to achieve a perceptual uniform chromaticity diagram. Even for this $u'v'$ diagram, it is only relative uniform for constant lightness. Therefore, using only diagram for color difference should be avoided. Chromaticity diagram with color has been used in all areas to illustrate the colors coverage by TV and display systems. It should be noted that those colors used are never the real colors as a chromacity coordinate is not enough to represent a color.

\[
\begin{align*}
x &= \frac{X}{X+Y+Z} \\
y &= \frac{Y}{X+Y+Z} \\
u' &= \frac{4 \times X}{X + 15 \times Y + 3 \times Z} \\
v' &= \frac{9 \times Y}{X + 15 \times Y + 3 \times Z}
\end{align*}
\]  

(2.5) (2.6)

2.2.4 Color Spaces

A color space is a coordinate system which is used to describe colors. Therefore, any tristimulus value is also called coordinate values in that color space. XYZ is the CIE tristimulus color space, as well as the CIE RGB. In addition to the tristimulus color spaces, CIE es-
established two opponent color spaces, close to lightness, chroma, hue spaces, CIELAB and CIELUV. A short summary of the equations is provided in Eq. 2.7 (CIELAB), Eq. 2.8 (CIELUV).

\[ \begin{align*}
X_n, Y_n, Z_n \text{ are the } XYZ \text{ of the white point used. The } X/X_n \text{ is used to model the chromatic adaptation mechanism. The nonlinear cube root compression is used to model the nonlinear response of the photoreceptors. } C_{ab}^*, h_{ab} \text{ are cylindrical representation of the same system. The parameters for CIELUV in Eq. 2.8 have the similar meanings.}
\end{align*} \]

\[ \begin{align*}
L^* &= 116 \left( \frac{Y}{Y_n} \right)^{1/3} - 16 \\
a^* &= 500 \left[ \left( \frac{X}{X_n} \right)^{1/3} - \left( \frac{Y}{Y_n} \right)^{1/3} \right] \\
b^* &= 200 \left[ \left( \frac{Y}{Y_n} \right)^{1/3} - \left( \frac{Z}{Z_n} \right)^{1/3} \right] \\
C_{ab}^* &= \sqrt{a^*^2 + b^*^2} \\
h_{ab} &= \tan^{-1} \left( \frac{b^*}{a^*} \right)
\end{align*} \] (2.7)

\[ \begin{align*}
L^* &= 116 \left( \frac{Y}{Y_n} \right)^{1/3} - 16 \\
u^* &= 13 \cdot L \cdot (u' - u'_n) \\
v^* &= 13 \cdot L \cdot (v' - v'_n) \\
C_{uv}^* &= \sqrt{u^*^2 + v^*^2} \\
h_{uv} &= \tan^{-1} \left( \frac{u^*}{v^*} \right)
\end{align*} \] (2.8)

In addition to the CIELAB, and CIELUV, there are many standard color spaces proposed by other organizations for specific applications. Especially for television and internet communication, a standard color space would facilitate the communication and the applications. sRGB, standard Red Green Blue, is the most well known standard color space [8]. This was
developed by HP and Microsoft in 1996. sRGB adopts the ITU-R BT 709 primaries and a
gamma close to 2.2. This gamma is close to the CRT display. Eq. 2.9 shows the conversion
from $XYZ$ (normalized) to $RGB_{linear}$, which will be compressed through Eq. 2.10. It should
be noted that the conversion from $XYZ$ to $RGB$ requires the input $XYZ$ should be that
under D65 illuminant.

$$
\begin{bmatrix}
R_{linear} \\
G_{linear} \\
B_{linear}
\end{bmatrix} =
\begin{bmatrix}
3.2406 & -1.5372 & -0.4986 \\
-0.9689 & 1.8758 & 0.0415 \\
0.0557 & -0.2040 & 1.0570
\end{bmatrix}
\begin{bmatrix}
X_{D65} \\
Y_{D65} \\
Z_{D65}
\end{bmatrix}
\tag{2.9}
$$

\begin{align*}
y &= 12.92 \times x (x \leq 0.0031308) \\
y &= 1.055 \times x^{1/2.4} - 0.055 (x > 0.0031308) \tag{2.10}
\end{align*}

### 2.2.5 Color Difference

Another important part of colorimetry is color difference. Color matching functions would
determine if two stimuli match with each other for an average observer, but it would not
indicate their difference if they do not match. A color difference formula was built on
CIELAB color space. First, color difference was defined by the Euclidean distance in CIELAB
color space, Eq. 2.11. It was found that the $\Delta E_{ab}^*$ does not have the uniformity across the
entire color space, especially for the high chroma area. The CIE proposed an improvement
for color difference in 1994, called $\Delta E_{94}^*$ Eq. 2.12. This formula implemented the chroma-
dependent parameters. For different applications, $k_L$, $k_C$, and $k_H$ can be adjusted. The
CIE also established some reference conditions, viewing condition, lighting condition and
the objects. The most recent CIE color difference was developed in 2000, called CIE color
difference 2000 [9]. Compared with $\Delta E_{94}^*$, a modification at the blue area was added due to
the poor performance of $\Delta E_{94}^*$ at that area. The full details can be found in [9].
\[
\Delta E_{ab}^* = \sqrt{\Delta L^*}^2 + \Delta a^*^2 + \Delta b^*^2
\]

(2.11)

\[
\Delta E_{94}^* = \sqrt{\left(\frac{\Delta L^*}{k_L \cdot S_L}\right)^2 + \left(\frac{\Delta C^*_{ab}}{k_C \cdot S_C}\right)^2 + \left(\frac{\Delta H^*_{ab}}{k_H \cdot S_H}\right)^2}
\]

\[
S_L = 1
\]

\[
S_C = 1 + 0.0045 \cdot C_{ab}^*
\]

\[
S_H = 1 + 0.0015 \cdot C_{ab}^*
\]

(2.12)

### 2.2.6 Metamerism

![Illustration of lighting metamerism](image)

Figure 2.5: Illustration of lighting metamerism. Different color masks over the lighting sources indicate different lights.

Metamerism is an interesting and important concept in color science. But it is also the most misunderstood concept. Also, this is more server for the wide color gamut (WCG) display. Therefore, it is necessary to introduce it here. Metamerism is an phenomenon where a pair of colors match with each other but mismatch under another condition. There are two types of metamerism: lighting metamerism and observer metamerism. Figure 2.5 illustrates the lighting metamerism. There are three important parts in the illustration:
Figure 2.6: Illustration of observer metamerism. Different color masks over observers/sensors indicate different observers/sensors.

lighting source, object, and observers/sensors. For lighting metamerism, the two objects would match under lighting A but mismatch under lighting B. For observer metamerism, the two objects would match for one observer/one type of sensors, but they mismatch for another observer/another type of sensor. Figure 2.6 illustrates the observer metamerism. Such a pair of objects are called metamers. Metamerism should be separated from corresponding color. The corresponding color is where one color under condition A looks the same for another color under condition B. The corresponding color is one color under one condition at one time. Metamerism is always about a pair of objects under different conditions.

The reason causing the matemerism is the spectral mismatch of objects. As introduced before, a color is described by a triple of tristimulus values. However, the tristimulus values are computed from the spectral color matching functions. Mathematically, it is possible that different spectral power distributions would result in the same tristimulus values. Also, the color matching functions are based on average data over a group of observers. Individual color matching functions have slight difference from the average data. Therefore, in observer metamerism, the variance over observers’ color matching functions would result in the mismatch. Asano and Fairchild published the most recent model about individual color matching functions, taking age, visual field, etc into consideration [10].
2.3 Color Appearance Model

Colorimetry was designed to help describe colors, and hence a better color reproduction. Those would be only valid under certain viewing condition, and lighting condition for the average observer. In the real practical realm, it is difficult to meet all the requirements from the CIE standards. Therefore several factors would affect the color appearance: background, surrounding conditions, the absolute luminance level. This motivated the development of color appearance models. Before introduction of color appearance model, we will briefly introduce some color appearance phenomena.

**Simultaneous contrast** is the most well known effect of background. Figure 2.7 shows an example. In figure 2.7 A, two identical circles under the same backgrounds look identical as they are physically the same. While in figure 2.7 B, the same two circles look different when one is under dark background and the other is under the bright background. The one under dark background looks brighter compared with the one in figure 2.7 A. The one under bright background looks darker compared with the one under the gray background in figure 2.7 A. So a dark background would induce the stimulus to appear lighter and a light background would induce the stimulus to appear darker. This is also valid for chromatic appearance. A red background would induce the stimulus to appear greener, vice verse. A yellow background would induce the stimulus to appear bluer and vice verse. This has been used a lot in art and design. More details can be found in Josef Albers’ book [11].

**Crispening** is an effect that the color difference between a pair of stimuli would increase when the background is close to the pair. A comprehensive study about the crispening effect and a prediction model was published in 1970 by Semmelroth [12]. Figure 2.8 illustrates an example. The pair of patches has obvious visual difference on the gray background. The same pair of patches looks almost the same over the black background and white background.

**Bezold-Brucke hue shift** is an effect that hue shifts with luminance level. It was assumed that hue can be specified by the wavelength of a monochromatic light. It was found
Figure 2.7: Example of simultaneous contrast. Two circles look the same under the same backgrounds (A). The two identical circles look different when under different backgrounds (B).
Figure 2.8: Example of crispening. The pair of patches on the three different backgrounds are the same.

that it changes when the luminance level changes.

**Abney effect** is a phenomenon that hue shift with colorimetric purity. It was assumed that white mixed with a monochromatic light should have constant hue as the monochromatic light. The colorimetric purity is the proportion of the monochromatic light over the mixed lights. In chromaticity diagram, the constant hue should be a line connecting the monochromatic light and the white. However, the contours of constant hue is not a straight line in chromaticity diagram.

**Hunt effect** was named in a Hunt study in 1952 on the light and dark adaptation. Hunt found that the perceived colorfulness increases with the luminance level of the stimulus. This highlights the importance of considering the absolute luminance level in color appearance.

**Stevens effect** is that the perceived contrast increases with absolute luminance level. Similar to Hunt effect, it shows the effect of absolute luminance level. Steven effect is about the contrast in lightness level, and Hunt effect is about chromatic contrast.

**Bartleson-Breneman** published equations predicting the effect from surrounding absolute luminance level. Perceived contrast was found to increase with an increasing surrounding
absolute luminance level. This is a key factor to consider in a complex practical situation.

Theoretically, CIELAB is a color appearance model, which includes a chromatic adaptation and a nonlinear compression. However, CIELAB does not take the background into consideration, hence not predicting simultaneous contrast, etc. There are several additional color spaces and color appearance models be introduced here.

**Nayatani Model:** Nayatani model takes in the chromaticity coordinates, which then are converted into the cone response. It also includes a complete output of colorfulness, brightness, lightness, chroma, etc. However, it does not account for the background effect, surrounding effect, rods’ response, etc. The computation of the model is fairly complex.

**Hunt Model:** is a very complex model counting background effect, surrounding effect, rods response. However, the model is so complex that it cannot include an inverse model.

**CIECAM97s:** is a simple version of Hunt Model. CIECAM97s includes background effect, surrounding effect, but not rods response. It also introduces a new chromatic adaptation degree $D$. $D$ is affected by surrounding effect, absolute luminance level. Also, the absolute luminance level, surrounding would affect the colorfulness, brightness etc.

**CIECAM02:** is an improved version of CIECAM97s. A major improvement is that the adaptation, von Kries, is a direct linear transformation of the tristimulus values instead of the chromaticity coordinates. Also, some small improvements of some constants were implemented.

**CAM02-UCS:** is a color space proposed by Luo [13]. The goal is a color appearance model and also a good fitting color difference model. It suggested a small change of the formula to have a better fitting in color difference data through optimization.

**CAM16:** is a color space proposed by Li, et al [14]. Mainly, this is used to correct the possible negative value from CIECAM02. Also, they proposed the corresponding uniform space based on CAM16.
2.4 Psychophysics

Psychophysics is a study of measuring/building the relationship between a physical stimulus and the sensation/perception evoked by the stimulus. Psychophysics provides most of the measured data for human color vision system knowledge. A very excellent recent book about application of psychophysics to image quality is written by Engeldrum [15]. In this section, we would briefly introduce some concepts and measure methodology about psychophysics.

Scale is important to understand and interpret a psychophysical experimental result. There are four important scales in psychophysics: nominal scale, ordinal scale, interval scale, and ratio scale. Nominal scale is that observers scale the stimuli by name, i.e. color names. So the result is just categories, there is no quantitative property. Ordinal scale is that observers scale stimuli by order for a certain criteria, like preference, colorfulness etc. The result would only be interpreted as rank order. For example, observers were asked to put the six images in order of individual’s preference. The average would only mean a certain order of preference for the group observers but not how much the first preference is over the second preference. Interval scale is that the difference between two scale points is comparable with each other but the absolute number has no meaning. For example, the difference between A and B is twice of that difference between C and D. Ratio scale is the same as interval scale with the absolute meaning. A, as a ratio scale, can be twice of B for certain property.

Usually the psychophysical experiments fall into two classes: threshold/small difference measurements; and large scaling experiments. For threshold, it is also called just noticeable difference (JND), which is the smallest difference to be perceived by a human. There are several typical methods for measuring JND: method of constant stimuli, method of adjustment, method of limits. The details of these methods will be not be discussed here. Here is a brief summary of these methods. In method of constant stimuli, several different intensity stimuli around the JND threshold should be chosen for observations. A psychometric curve can be derived based on the “seen” probability against the physical property of these chosen
stimuli, i.e. luminance level. Another way of method of constant is measuring the JND over a certain intensity stimulus. For this way, a pair will be presented to observers. The pair will always have one as this certain intensity stimulus and the other stimulus is around the threshold above or below the constant stimulus. A slight different probability psychometric curve function will be derived to fit the observers’ data. Details can be found in Engeldrum’s book. In **method of adjustment**, observers are given the access to adjust the intensity of the stimulus to the level that it reaches their own thresholds. The average of these individual thresholds can be computed as the group threshold. In **method of limits**, observers are asked to make judgement over predefined steps in a ascending or descending order until they have a different answer compared with the previous level. The intensity of the stimulus then changes order, from ascending to descending or descending to ascending. The average of the pivots, where the changes happen, would be considered as individual threshold. For large scale measurement, usually there will be anchor point / anchor points provided. Observers then would decide a scale estimation of a stimulus based on the anchor points. This is used often in quality assessment. More details can also be found in *Engeldrum’s* book [15].
Chapter 3

High Dynamic Range Display and Related Work

The absolute luminance of the world can go up to $10^8 \, cd/m^2$ as under direct sunlight, and down to $10^{-6} \, cd/m^2$ at night. The human vision system has the ability to adapt to such a large range. However, this should be separated from simultaneous dynamic range, which refers to the range of the scene shown simultaneously. For a reflective matte object under diffuse lighting, the dynamic range is limited to around 20:1. A direct lighting would increase the dynamic range to 100:1. If the material has specular highlights, the dynamic range would increase dramatically. Of course, direct lighting inside the scene would also significantly boost the dynamic range. The standard dynamic range (SDR) display has a peak luminance of 100 $cd/m^2$ and a black level of 0.5 $cd/m^2$ [16]. However, this was far from sufficient to provide a state-of-the-art rendering. Later on, the commercial display would achieve a peak luminance level of 200 – 300 $cd/m^2$ and a black level of 0.5 $cd/m^2$. In order to display a high dynamic range image on the limited dynamic range displays, tone mapping operators (TMO) were developed to provide a better usage of the display in showing high dynamic range images. TMO is an algorithm optimizing the usage of SDR display for HDR contents. However, it is not enough to provide a state-of-the-art experience for customers.
Recently, high dynamic range (HDR) displays have gained popularity. A recent HDR display can achieve a peak luminance at least $600 - 700 \text{ cd/m}^2$, and a black level no more than 0.01 $\text{cd/m}^2$. Along with the development of HDR displays, more visual studies on HDR have been published. In this chapter, we will introduce HDR displays, HDR color models, and some HDR related works.

### 3.1 High Dynamic Display

Before introducing HDR displays, the configuration of LCD will be briefly explained. Liquid Crystal Display is a type of display using crystal array to modulate the light intensity, which is usually from back of the display. A typical LCD display would have a back-light on the edge of the panel, which will illuminating the panel before a diffuser layer. The diffuser layer would make the light more evenly across the whole panel. The liquid crystal would be used to modulate the intensity of the light. One big drawback of the LCD is the back-light configuration. The back-light would be full on all the time, therefore it will be strong internal scattering especially if some pixels are bright and some are dim. Also, the modulation ability of the liquid crystal has some physical limitation. Due to the two constrains, there is a limitation about the dynamic range of such LCDs.

Seetzen from Sunnybrook Tech published an interesting methodology of generating a high dynamic range display. They used a projector combined with a layer of LCD to modulate the lights. Figure 3.1 illustrates the configuration.

The projector can provide a contrast ratio $c_1$, peak luminance to minimal luminance. The LCD could modulate the input lighting at a $c_2$ contrast. The combined system can demonstrate a contrast of $c_1 \times c_2$. The projector has a dynamic range of 800:1, and the LCD has a dynamic range of 300:1 as a typical LCD. Theoretically, this configuration can reach 240,000:1. The actual contrast of this display was measured as 54,000:1. The actual peak luminance of this HDR display was 2700 $\text{cd/m}^2$ with a black level of 0.05 $\text{cd/m}^2$ [17]. There

26
Figure 3.1: Schematic digital light projector (up) and the real configuration (bottom). [17].
were two drawbacks of this configuration: the high power consumption and the complex algorithm to compensate for blurring. A liquid crystal panel has only a transmittance of 5% - 8%, which yielded a very high power consumption. Moreover, due to the uniformity and low resolution of the projector, the implementation of diffuser would blur the image, which could be compensated algorithmically given a good measurement of the point spread function (PSF) of the projector. Wanat [18] presented a detailed study of the physical limitations of the projector based HDR display. It also included a psychophysical experiment about the perceptual limit in HDR displays. We will discuss it later.

In order to overcome the high power consumption of the projector-based HDR display, they proposed another type of HDR display: light emitting diodes (LED) based HDR display. These diodes were programmable and had lower power consumption compared to the projector. Also the LED provided a higher contrast ratio than the projector. The LED based HDR display achieved $8500\, cd/m^2$ peak luminance and $0.03\, cd/m^2$ black level. Theoretically, multi-layer of LCDs could achieve much larger contrast ratio but with much high power consumption. Also, multi-layer would require much more complex construction. Usually, two-layer LCDs would only be used in the laboratory settings.

**OLED**, organic light-emitting-diodes, high dynamic range displays have been growing quickly in recent decades. The first OLED device was built in Eastman Kodak in 1987 [19]. Due to its advantage in minimal luminance level, a deeper black, OLED could achieve a higher contrast ratio. The OLED display has a unique character that the organic material emits light directly while a traditional LCD modulates back-lit lights through the liquid crystal panel. The liquid crystal has very low transmittance 5% - 8%. Another disadvantage of LCD is the back light. The LCD is lit by the whole panel, which usually has some light leak. The maximum dynamic range contrast of a LCD is around 1000:1. This is far from a satisfying visual experience. Recently, a local dimming technology was used to improve the blackness of LCD. Local dimming technology uses multi-array back lights, so each area can be turned on or off separately [20, 21]. But the more arrays the back lights has, the
more complex the system is. Therefore, most manufacturers would keep a balance between the deep black and the complexity of the system. While for OLED displays, each pixel can be turned off/modulated separately. Therefore, OLED can achieve a much deeper dark level. OLED can easily reach a black level of 0.001 nits. Additionally, OLED has a few more advantages: light weight, flexibility, better power efficiency, and faster response time. OLED displays can be fabricated on flexible plastic substrates, which makes OLED light weight and flexible. Also, almost 100% transmittance of OLED gives OLED high power efficiency. Moreover, OLED has a much faster response time than LCD. OLED theoretically can achieve a much higher refresh rate, up to 1000 times that of LCD. OLED has several disadvantages: life time, color balance, water damage, ambient light sensitivity. Firstly, OLED has a short life time: around 14,000 hours for the maximum peak luminance to decrease to half. However, LCDs usually have lifetimes of 25,000 - 40,000 hours. This is mainly due to the property of the organic material used. Another related drawback is the color balance. The different degradation speeds of the RGB channels would cause the color balance to change with time. The other well-known disadvantage is water damage: a small amount of water or water vapor can cause severe damage to OLEDs. Therefore, the OLED device requires very advanced sealing technology. Finally, the cathode in OLED has high reflectivity. Therefore, ambient light would deteriorate the black advantage of OLED.

The British Broadcasting Company (BBC) has published two articles about the adjustment for using HDR displays under domestic viewing environments up to 500 cd/m² ambient illuminations [22, 23].

To achieve a high dynamic range, LCDs could lift their peak luminance by increasing the power consumption within the hardware limit, while OLEDs can achieve the same contrast ratio through their very low black level. OLEDs cannot achieve very high peak luminance due to the material property. Moreover, usually an OLED display can only achieve its peak luminance for a small area, but not for the full screen. Kwon’s measurement showed that the peak luminance starts to decrease when the testing window size reached 10% [25]. Tian’s
Figure 3.2: Peak luminance VS window size. Absolute luminance level is on the top, and normalized luminance is on the bottom [24].
Figure 3.3: Peak luminance heat map for background colors and window size [25].

Recent publication demonstrated the similar result that 10% window size is a break point where the peak luminance started to decrease [24]. Figure 3.2 is the measurement result from [24]. W channel in the plot is the data of the white channel. The OLED display usually adopts RGBW configuration to achieve a higher peak luminance. The normalized plot of RGBW channels showed that the white channel started to decrease at 10% window size. The blue channel started to decrease at around 30% window size. The red channel and green channel also started to decrease at around 10%. Moreover, due to the different degradation rates against the window size, the peak white would also shift if the window size reached 10%. The background of the display also affected the peak luminance of the testing window. Figure 3.3 is the plot of peak luminance heat map for different window sizes and background colors [25]. It can be found that when the background reached 150 (in 8 bit), the peak luminance was clearly affected.

In addition to the high dynamic range, wide color gamut is another major advantage of the OLED display compared with the LCD. Usually, the color gamut of a display was defined by the chromaticity diagram coverage of the primaries of a display. There are three
well-known widely used primary sets: Rec 709, DCI/P3 and Rec 2020. The color gamuts of the Rec.709, DCI/P3, and Rec. 2020 in \( xy \) chromaticity diagram and \( u’v’ \) chromaticity diagram are plotted in Figure 3.4. Again, this does not represent the display’s color ability. Also the difference between any two points on the chromaticity diagram does not necessarily represent their color difference. This is also the reason that the chromaticity diagram of the spectrum (spectrum locus) is plotted in midgray instead of any color. A traditional LCD would reach 90% or 100% coverage of the Rec 709. OLED can easily reach almost 100% coverage of DCI/P3, which has 26% larger chromaticity area coverage than Rec.709 in \( xy \) chromaticity diagram. Many Apple products displays have already achieved 100% DCI/P3. In order to improve the color gamut of the LCD display, a new technology, quantum dots (QD), was developed and applied. The QD was directly applied onto the light emitting diodes (LED) [26]. It works like a filter, which could narrow down the spectral bandwidth of the LED. Luo showed several measurements of QD LEDs with wide color gamut [27]. Zhu proposed a simulation of reaching the Rec. 2020 using QD for LEDs [28].

![Figure 3.4: Primaries of Rec. 709 (smallest triangle), DCI/P3 (mid-size trangle), and Rec. 2020 (largest triangle) in \( xy \) chromaticity diagram (left) and in \( u’v’ \) chromaticity diagram.](image-url)
3.2 HDR Color Appearance Models

3.2.1 hdrCIELAB & hdrIPT

In order to build a color space for high dynamic range, *Chen* conducted two experiments to explore the lightness below and above diffuse white [29, 30]. A short summary will be presented below.

Figure 3.5 (left) shows the setup of the experiment I. In experiment I, three 4.8° patches were surrounded by a 1 inch midgray (\(L^* = 50\)) background and a 2 inch wide paper white. This setup was designed to help the observer use the paper white as reference white/diffuse white (\(L^* = 100\)). Observers were asked to adjust one of the patches to make the difference between center and right equal to the difference between left and center. This is called the method of partition scaling. This experiment consisted of two parts: lightness above diffuse white (called \(SL > 100\)) and lightness below diffuse white (\(SL < 100\)). It should be noted that the reference white was the same as the outside 2 inch width paper white.

Figure 3.5 (right) illustrates the setup of experiment II. In this experiment, a key difference is that the patch size was 2°. This experiment also consisted of the measurements for \(L^* > 100\) and \(L^* < 100\). A slight different methodology was used for measuring \(L^* < 100\). For \(L^* < 100\), sufficient prepared patches with \(L^*\) from 0.5 to 90 were presented to the observers to make several equal visual difference steps from imagined black to the diffuse white. Individual result of \(L^* < 100\) was presented (red box in Figure 3.5) for observers who conducted \(L^* > 100\). The three patches, dark, diffuse white and a tunable bright patch, were on the reference white background (green box in Figure 3.5). The observers were asked to adjust the bright patch to equalize the difference between center and right to the difference between left and center, the same methodology as experiment I.

Based on the data from the two experiments, two color spaces were developed for the lightness above diffuse white: hdr-CIELAB and hdr-IPT [31, 32]. Mainly, a different nonlinear compression function was developed to replace that in CIELAB. Initially, it was reported
Figure 3.5: Set up of experiment I (left) and experiment II (right) from Chen’s work [30].

by Fairchild and Wyble [31]. An improved version was reported by Fairchild [32]. A short summary of the improved version is introduced here.

**hdr-CIELAB:** Eq. 3.1 was the derived equation for computing hdr-CIELAB, replacing the \( f \) nonlinear regression. Eq. 3.2 is the optimized nonlinear compression based on the experimental data. The authors also proposed the effect from the surrounding \( sf \), and luminance level \( lf \) in Eq. 3.3. The \( sf \) was used to account for the Bartleson-Breneman effect. The \( lf \) was to used to account for Stevens effect. \( Y_s \) was the relative luminance level of the surrounding, and \( Y_{abs} \) was the absolute luminance level of the scene diffuse white in cd/m\(^2\). It should be noted that \( f(\omega) \) Eq. 3.2 is only valid for \( \omega \) from 0 to 4 according to the statement from [32].

\[
\begin{align*}
L_{hdr} &= f(Y/Y_n) \\
a_{hdr} &= 5 \cdot [f(X/X_n) - f(Y/Y_n)] \\
b_{hdr} &= 2 \cdot [f(Y/Y_n) - f(Z/Z_n)] \\
C_{hdr} &= \sqrt{(a_{hdr})^2 + (b_{hdr})^2} \\
b_{hdr} &= \tan^{-1} \left( \frac{b_{hdr}}{a_{hdr}} \right)
\end{align*}
\]
\[ f(\omega) = 247 \times \frac{\omega^\epsilon}{\omega^\epsilon + 2^\epsilon} + 0.02 \]  

\[ \epsilon = 0.58 / (sf \times lf) \]

\[ sf = 1.25 - 0.25 \times (Y_s/0.184); \ (0 \leq Y_s \leq 1) \]  

\[ lf = \log(318)/\log(Y_{abs}) \]  

hdr-IPT was derived under the same constraints and procedure. The nonlinear compression of \( LMS \), \( f \), in Eq. 3.6 was replaced by a new nonlinear compression, Eq. 3.4, derived from the experiments. Eq. 3.5 was used to counter the surrounding effect \( sf \) and the absolute luminance level effect \( lf \). The \( sf \), \( fl \), \( Y_s \) and \( Y_{abs} \) are the same meanings as in hdr-CIELab model.

\[ f(\omega) = 246 \times \frac{\omega^\epsilon}{\omega^\epsilon + 2^\epsilon} + 0.02 \]  

\[ \epsilon = 0.59 / (sf \times lf) \]

\[ sf = 1.25 - 0.25 \times (Y_s/0.184); \ (0 \leq Y_s \leq 1) \]  

\[ lf = \log(318)/\log(Y_{abs}) \]
\[
\begin{bmatrix}
L \\
M \\
S
\end{bmatrix} =
\begin{bmatrix}
0.4002 & 0.7075 & -0.0807 \\
-0.2280 & 1.1500 & 0.0612 \\
0.0 & 0.0 & 0.9184
\end{bmatrix} \cdot
\begin{bmatrix}
X_{D65} \\
Y_{D65} \\
Z_{D65}
\end{bmatrix}
\]

\[
L' = f(L); \quad L \geq 0
\]

\[
L' = -f(-L); \quad L < 0
\]  \hspace{1cm} (3.6)

\[
M' = f(M); \quad M \geq 0
\]

\[
M' = -f(-M); \quad M < 0
\]

\[
S' = f(S); \quad S \geq 0
\]

\[
S' = -f(-S); \quad S < 0
\]

\[
\begin{bmatrix}
I_{hdr} \\
P_{hdr} \\
T_{hdr}
\end{bmatrix} =
\begin{bmatrix}
0.40 & 0.4 & 0.2 \\
0.4550 & -4.851 & 0.396 \\
0.8056 & 0.3572 & -1.1628
\end{bmatrix} \cdot
\begin{bmatrix}
L'_{D65} \\
M'_{D65} \\
S'_{D65}
\end{bmatrix}
\]  \hspace{1cm} (3.7)

\[
C_{hdr-IPT} = \sqrt{(P_{hdr})^2 + (T_{hdr})^2}
\]

\[
h_{hdr-IPT} = \arctan^{-1}\left(\frac{T_{hdr}}{P_{hdr}}\right)
\]

As is known, CIELAB has its own nonlinear compression function. There is a simple way of applying CIELAB in high dynamic range above diffuse white by extending the limit of $L$ above 100. The performance comparison between the extended CIELAB, hdr-CIELAB, and hdr-IPT can be found in detail in [30, 32]. Analysis of the performance on Munsell colors showed that no significant difference was found between hdr-CIELAB, hdr-IPT and the extended CIELAB, IPT [32]. An optimization of the nonlinear compression was conducted to improve the hdr-CIELAB and hdr-IPT in predicting lightness scales above diffuse white. The optimized compression function for hdr-CIELAB follows Eq. 3.8 and that for hdr-IPT
is in Eq. 3.9. The optimized two compressions demonstrated similar performances with the CIELAB and IPT compression for relative lightness from 0 to 6. Therefore, extending CIELAB and IPT directly is not a bad option based on evaluation of these two experimental data.

\[ f(\omega) = 253 \times \omega^{0.61} / (\omega^{0.61} + 2^{0.61}) + \omega^{1.88} \]  

\[ f(\omega) = 261 \times \omega^{0.65} / (\omega^{0.65} + 2^{0.65}) + \omega^{2.09} \]  

\[ (3.8) \]

\[ (3.9) \]

3.2.2 PQ & ICtCp

As introduced before, an image or color can be decomposed into the luminance channel and the chromatic channel. There are several different characteristics for the luminance channel and the chromatic channels. Firstly, the luminance channel provides much more information than the chromatic channels. Also, human exhibit different contrast sensitivity functions (CSFs) to the luminance channel and the chromatic channels. The chromatic channels have their peak contrast sensitivity at 0 spatial frequency but the CSF of the luminance channel has the peak around 2-5 cycle/degree. Also the CSF of the luminance channel has a much higher cutoff frequency than that of the chromatic channels. The CSF also varies with the absolute luminance level. The CSF, as a function of luminance level, is usually used as a criteria to evaluate the quantization effect of image signal encoding. Image data is encoded in bit data, i.e. Rec 709 has standard of 8 bit for each color channel. The CSF can be used to determine if the encoding would result in any possible banding. If the contrast between the two consecutive levels in the encoding system exceeds the maximum sensitivity of the CSF, there is a possibility that the banding would be perceived by observers. Barten published a
comprehensive study of the CSF formula for human eyes [33]. The publication showed that the CSF as a function of spatial frequency is also affected/modulated by luminance level, visual size, pattern orientation and surrounding luminance level. Figure 3.6 shows the CSF as a function of spatial frequency for different luminance levels (A) and different visual sizes (B).

![Figure 3.6: CSF as a function of spatial frequency for different absolute luminance levels (A) and different visual sizes (B) [33].](image)

Miller developed a nonlinear encoding method based on Barten’s formula for high dynamic range systems [34]. This nonlinear encoding is called Perceptual Quantizer (PQ). Miller chose 40° visual size for the maximum sensitivity. They calculated the peak CSF as a function of the absolute luminance. That was used to develop and evaluate the nonlinear compression algorithm. Figure 3.7 shows the contrast of 12-bit PQ encoding system for different peak luminances. It showed that 12-bit PQ can reach the contrast thresholds from Barten’s model. That ensured no banding effect from the 12-bit PQ encoding algorithm. It was found that PQ showed significant better performance than Rec. 1886 (gamma) encoding method. A gamma nonlinear encoding seems to waste too much bit depth for the bright end but not enough in the dark end. Miller conducted a visual test to validate the PQ system. The visual test verified that 10-bit or 11-bit is always enough to eliminate any visual quantization artifacts with PQ. Artifacts from gamma at low luminance level was always perceived.
by observers even with high bit depth.

![Graph](image)  

**Figure 3.7:** Comparison of different encoding systems and with the Barten’s sensitivity model [34].

Based on the PQ encoding system, a new color space was developed by Dolby Vision, ICtCp. The formula can be found in a white paper [35, 36]. A short summary is presented here Eq 3.10 - 3.12. Eq 3.10 is the conversion from RGB linear in BT. 2020. The result of Eq 3.10 then is encoded through the PQ⁻¹, which is standardized and recommended in ITU-R [37]. The PQ in [37] is a 12-bit nonlinear encoding system, peaking at 10,000 nits and minimal at 0.005 nits. This nonlinear L’M’S’ is converted into IC₇C₇ through Eq 3.12, which is similar with the conversion in IPT spaces [38] for better hue linearity. It should be noted that IC₇C₇ is more of an encoding system to avoid any possible artificial banding effect rather than a HDR color appearance model. Color appearance model should predict some visual effect, i.e. simultaneous contrast, Stevens effect, Hunt effect. IC₇C₇ obviously does not include the adaptation mechanism, either achromatic adaptation or chromatic adaptation. Therefore, IC₇C₇ cannot predict most visual effects. However, IC₇C₇ has some similarities
with Brightness-Colorfulness-Hue space in terms of the axes meaning.

$$\begin{bmatrix} L \\ M \\ S \end{bmatrix} = \frac{1}{4096} \begin{bmatrix} 1688 & 2146 & 262 \\ 683 & 2951 & 462 \\ 99 & 309 & 3688 \end{bmatrix} \begin{bmatrix} R'_{BT2020} \\ G'_{BT2020} \\ B'_{BT2020} \end{bmatrix} \tag{3.10}$$

$$L'M'S' = \text{EOTF}^{-1}_{PQ}(LMS) \tag{3.11}$$

$$\begin{bmatrix} I \\ C_T \\ C_P \end{bmatrix} = \frac{1}{4096} \begin{bmatrix} 2048 & 2048 & 0 \\ 6610 & -13613 & 7003 \\ 17933 & -17390 & -543 \end{bmatrix} \begin{bmatrix} L' \\ M' \\ S' \end{bmatrix} \tag{3.12}$$

Figure 3.8: Example of testing example from [39].

An interesting work on color difference for HDR WCG imagery was presented by Pieri [39]. As is known, the existing color difference data were measured between around 15 cd/m² and 100 cd/m². Pieri conducted a color difference experiment at 0.1, 25 and 1000 cd/m². Four alternative forced choice was used as measuring methodology to improve naive observers’ performance. The background of the testing patch was a 1/f noise pattern, which has the same average luminance level of the testing patches. Figure 3.8 is a screenshot from
Figure 3.9: Results of the color difference predictions [39].
the example from [39]. The authors compared the proposed color difference in IC<sub>T</sub>C<sub>P</sub> as in Eq. 3.13 and Δ<sub>L</sub><sup>*</sup>u<sup>*</sup>v<sup>*</sup>, ΔE94, ΔE00, ΔCAM02 − UCS. The results of the comparison between ΔE00 and ΔIC<sub>T</sub>C<sub>P</sub> is shown in Figure 3.9. The plots show very similar performance between ΔE00 and ΔIC<sub>T</sub>C<sub>P</sub> for 25 cd/m<sup>2</sup> and 1000 cd/m<sup>2</sup>, but significant better for ΔIC<sub>T</sub>C<sub>P</sub> than ΔE00 for 0.1 cd/m<sup>2</sup>. It raised up a question about how the calculations were made in converting the XYZ to CIELAB, which was not explained. The choice of white point is very important in this conversion. Since the experiment included a full screen of the same luminance level (the same average background and the test patches), it will be tricky to choose a suitable diffuse white to test ΔE00. Also, for 0.1 cd/m<sup>2</sup>, the sensitivity of the rods is more important than that of the cones. As is known, the rods is luminance channel only. Therefore, was the measurement of changing chromatic direction really necessary? Moreover, would that affect observers’ visual experience in a real image or video? This experiment demonstrated some explorations of the color difference beyond the traditional luminance range but there are some questions about the details and the necessity of testing color differences around the rods’ sensitivity level.

\[ \Delta IC_TCP = 720 \times \sqrt{(\Delta I)^2 + 0.25 \times (\Delta CT)^2 + (\Delta CP)^2} \]  

(3.13)

3.2.3 $J_z a_z b_z$

Safdar etc, published several articles about developing a color space for HDR WCG imagery/signals. The authors firstly attempted to investigate the performance of five color spaces, CIELAB, CIELUV, CAM16-UCS, IC<sub>T</sub>C<sub>P</sub>, and ICaCb, in local uniformity, global uniformity, hue linearity, convergence of hue lines, neutral point error, HDR support, WCG support, and computation complexity [40]. It is an interesting study about several characteristics of the popular color spaces. For convergence of hue lines, even if the hue lines do not have a very good convergence, it does not necessary mean that would be perceived by observers since all those colors are very neutral grays. Moreover, HDR and WCG support is
also more of the coding system than the property of the color spaces. For example, CIELAB was not designed to support above diffuse white levels but that does not exclude the possible good performance of CIELAB in extended diffuse white levels as introduced before. The authors published a new color space specifically for HDR and WCG applications, an optimized color space $J_z a_z b_z$ [41]. The new color space is an optimized result of the ICaCb with taking lightness scaling above diffuse white from Chen [29]. Eq. 3.14 - 3.21 are the mathematical model of $J_z a_z b_z$. $J_z a_z b_z$ also use similar transforms to IPT for a better hue linearity and a similar compression of color appearance. The $J$ is an modification of $I_z$, which was aiming to better predict the extended lightness. All the constant values are listed in Table 3.1.

\[
\begin{bmatrix}
X'_{D65} \\
Y'_{D65}
\end{bmatrix} = \begin{bmatrix}
b * X_{D65} \\
g * Y_{D65}
\end{bmatrix} - \begin{bmatrix}
(b - 1) * Z_{D65} \\
(g - 1) * X_{D65}
\end{bmatrix}
\] (3.14)

\[
\begin{bmatrix}
L \\
M \\
S
\end{bmatrix} = \begin{bmatrix}
0.414789 & 0.57999 & 0.0146 \\
-0.20151 & 1.1206 & 0.0531 \\
-0.0166 & 0.2648 & 0.6685
\end{bmatrix} \begin{bmatrix}
X'_{D65} \\
Y'_{D65} \\
Z_{D65}
\end{bmatrix}
\] (3.15)

\[
\{L' M' S'\} = \{\frac{c_1 + c_2 * (\{L M S\}/10000)^n}{1 + c_3 * (\{L M S\}/10000)^n}\}^p
\] (3.16)

\[
\begin{bmatrix}
I_z \\
a_z \\
b_z
\end{bmatrix} = \begin{bmatrix}
0.5 & 0.5 & 0 \\
3.524 & -4.066 & 0.5427 \\
0.199 & 1.0968 & -1.2958
\end{bmatrix} \begin{bmatrix}
L' \\
M' \\
S'
\end{bmatrix}
\] (3.17)

\[
J_z = \frac{(1 + d) * I_z}{1 + d * I_z} - d_0
\] (3.18)

\[
C_z = \sqrt{(a_z)^2 + (b_z)^2}
\] (3.19)
<table>
<thead>
<tr>
<th>parameter</th>
<th>value</th>
</tr>
</thead>
<tbody>
<tr>
<td>b</td>
<td>1.15</td>
</tr>
<tr>
<td>g</td>
<td>0.66</td>
</tr>
<tr>
<td>c₁</td>
<td>$3424/2^{12}$</td>
</tr>
<tr>
<td>c₂</td>
<td>$2413/2^7$</td>
</tr>
<tr>
<td>c₃</td>
<td>$2392/2^7$</td>
</tr>
<tr>
<td>n</td>
<td>$2610/2^{14}$</td>
</tr>
<tr>
<td>p</td>
<td>$1.7 \times 2523/2^5$</td>
</tr>
<tr>
<td>d</td>
<td>-0.56</td>
</tr>
<tr>
<td>$d₀$</td>
<td>$1.629 \times 10^{-11}$</td>
</tr>
</tbody>
</table>

Table 3.1: Constant parameters for $J_zazb_z$.

\[
h_z = \arctan\left(\frac{b_z}{a_z}\right)
\]

\[
\Delta E_z = \sqrt{\left(\Delta J_z\right)^2 + \left(\Delta C_z\right)^2 + \left(\Delta H_z\right)^2}
\]

3.3 Summary

In this chapter, we introduced some configurations of the HDR displays from the projector based to the new OLED HDR display. Also, the advantages and disadvantages of the OLED HDR and LCD HDR displays were explained. Moreover, the LCD HDR display mostly utilizes the local dimming technology, where the whole panel can be manipulated locally but not at pixel level. Additionally, some characterization work of HDR displays was summarized. However, it should be noted that due to some power limitation, there is no general colorimetric characterization for HDR displays, which are different from the typical
SDR display. Therefore, the characterization of HDR display will be a challenging topics.

Also, some related work about the HDR color space, and HDR encoding system were introduced. The PQ encoding system and some visual experiments were introduced and explained. The very original experiment about lightness scaling above diffuse white was summarized. Two HDR color spaces, $hdrCIELAB$ and $hdrIPT$, were derived based on the experiments’ data. An optimization of local uniformity, global uniformity, hue linearity was published for a HDR and WCG color space, $J_z a_z b_z$. However, obviously the color appearance model is not comprehensive yet. A large amount of research work is required to collect more experimental data, which will provide the visual data for HDR appearance model. Additionally, the fundamental difference between SDR and HDR should be explored more, as well as the effect from the wide color gamut.
Chapter 4

Color Gamut Volume

4.1 Introduction

Color is a visual sensation triggered by the photons reaching the retina and processed through our brain. There are many important attributes used to describe such a sensation: hue, lightness, brightness, saturation, chroma, etc. Hunt provided a comprehensive explanation about this terminology in 1978 [42] and it is standardized in the Commission Internationale de l’Eclairage (CIE) International Lighting Vocabulary [43]. Matching of some of these color attributes is often a goal of color reproduction systems. In color reproduction, color gamut sometimes becomes poor-defined terminology used to describe color range of a device or image content. Gamut, as a fundamental concept, is defined as a complete range of some variable in any give number of dimensions. When applied to colors, color gamut is used to describe the full range of colors, either in terms of color specification (e.g. CIE $u'v'$) or in terms of color appearance (e.g. CIELAB). The CIE has a defined color gamut as “volume, area, or solid in a colour space, consisting of all those colours that are either: (a) present in a specific scene, artwork, photograph, photomechanical, or other reproduction; (b) capable of being created using a particular output device and/or medium” [44]. The most well-known so-called color gamut is the chromaticity gamut area of a display in $u'v'$ or $xy$ chromaticity
diagram. In the recent decades, the 2D gamut in \(xy\) and \(u'v'\) chromaticity diagrams has been widely utilized by display manufacturers to characterize the display’s color reproduction ability. However, it is well known that at least a three-dimensional color space is necessary to describe a color precisely. This is also typically how color is encoded/recorded. Therefore, the 2D gamut is only an illustration of part of color rendition capabilities, but inappropriate for characterizing display’s color appearance reproduction. Another interesting usage of the 2D gamut is that the \(u'v'\) is generally more favored than the \(xy\) by the manufacturers due to it being slightly more perceptually uniform for threshold discrimination despite the fact that chromaticity diagrams do not describe perceptual appearance. This is based on CIE \(u'v'\) having better uniformity performance based on MacAdam’s experiment on the visual sensitivity to color difference in daylight experiments back in 1942, based on matching variability [45]. The experimental data demonstrated the better color threshold uniformity in CIE \(u'v'\). However, the experiment was based on one observer with the colors on constant luminance level. Obviously, this is not sufficient as a strong evidence to support the superiority of CIE \(u'v'\) over CIE \(xy\) as an approximation to the 3D color spaces, especially when neither describes color appearance directly. While the \(u'v'\) or \(xy\) chromaticy diagram can be very useful in illustrating color stimulus capability in a convenient way, the lack of the higher dimensions of appearance and strong evidence of good correlation between the 2D gamut and 3D color space limit the possibility of approximating 3D color volume using 2D chromaticity gamuts. Moreover, the area coverage by the primary of display on the chromaticity diagram is either an interval scale or a linear scale, i.e. the area coverage of one display can be said to be larger than the other but statements such as “30% larger or smaller” are meaningless in terms of color appearance.

A minimum of three dimensions are required to describe color. Thus, color gamut volume in three dimensional space is the meaningful method to describe device color reproduction capability. Hypothetically, the color gamut volume can be computed in any color space, e.g. sRGB, XYZ, Yxy, etc. For example, a 3D color gamut volume of a display can be
computed in XYZ as the mathematical volume occupied by all possible reproduced stimuli of a display. However, such spaces are more of the encoding of unrelated color stimuli, defined as perceived color from area or object in isolation from other stimuli by Hunt [42], since these color spaces do not include the two important steps in our visual system, adaptation and response compression. It is well known that the adaptation allows our visual system to perform perceptually over a wide luminance range from at least 0.01 cd/m$^2$ to $10^8$ cd/m$^2$ [46]. Response compression is another process, which is a nonlinear conversion from the physical linear signal into our visual sensation, that functionally expands the dynamic range of human perception. For example, the well-known Weber-Fechner law, a simplification of actual human performance, is modeling such compression. Color appearance models, as proposed and tested by CIE Technical Committee 1-34, are required to have predictive correlates of at least the relative appearance attributes of lightness, chroma, and hue [47]. CIELAB is the widely used and well-known example of such a color appearance space though its original goal was supra-threshold color-difference uniformity. CIELAB includes the very simple chromatic adaptation by scaling the X, Y, Z by the XYZ of the diffuse white and an approximate 1/2.3 power-function compression. However, CIELAB does not take the other attributes of the viewing environment and background, such as absolute luminance level, into consideration. CIECAM02 is CIE-recommended color appearance model for more complex viewing situations, and it predicts correlates of brightness ($Q$), colorfulness ($M$), and saturation ($s$) in addition to the lightness ($J$), chroma ($C$), and hue ($h$) [48]. In the years since adoption of CIECAM02, there are several new proposed color appearance models with some modifications and improvements, e.g. CIECAM02-UCS [49], CAM16-UCS [14], CAM16-UCS [14]. CIECAM02-USC is an adjustment of the CIECAM02 aiming at reconciling color appearance scales with supra-threshold small color difference equations, noting that both perceptual phenomena cannot be predicted using the same perceptual scales. CAM16 includes a modification to solve occasional inconvenient negative values in CIECAM02 through a simple adjustment to the chromatic adaptation transform. CAM16-UCS is similar to CIECAM02-
UCS but based on CAM16. There is no significant change in perceptual predictions of these new color spaces from the current CIE recommendation, CIECAM02. Therefore, color gamut volume computed in CIELAB and CIECAM02 are representations. Masaoka and Nishida reported an interesting finding about the RGB-primary display’s 3D color gamut volume in CIELAB and JCh in CIECAM02 [50]. That work demonstrated a high correlation between the 3D color gamut volume expressed in CIELAB, CIELUV, JCh and the primary area coverage in the CIE \( xy \) chromaticity diagram, better than achieved using the CIE \( u'v' \) chromaticity diagram. Later, Masaoka further analyzed the causes of these correlations and showed that it was due to the relative weighting of various chromaticity regions with respect to the display gamut volume in those regions [51]. Masaoka followed up that work by illustrating the importance of 3D color gamut volume in CIELAB for the comparison of High Dynamic Range (HDR) - Wide Color Gamut (WCG) displays [52]. In particular, this is critical when comparing RGB displays with RGBW displays or when comparing displays with disparate peak luminance levels or ratios between peak luminance and rendered diffuse white luminance.

Along with the development and widespread use of the HDR WCG displays, there have been color spaces proposed specifically for HDR WCG. Fairchild and Chen proposed and evaluated hdr-LAB and hdr-IPT color spaces based on experimental data of lightness above the diffuse white level [32]. The extended CIELAB still showed a reasonable performance for predicting the lightness above the diffuse white. Recently, there are two interesting proposed color spaces for usage in HDR WCG applications, ICtCp [53, 54] and \( J_2a_2b_2 \) [41]. ICtCp was developed on the perceptual quantizer (PQ) nonlinear function recommended by Society of Motion Picture & Television Engineers (SMPTE) [55]. The PQ is proposed based on Barten’s Contrast Sensitivity Function (CSF) model [33] to avoid quantization artifacts. \( J_2a_2b_2 \) utilizes the similar structure as ICtCp with a further optimization of some parameters based on a comprehensive experimental data. Both color spaces showed reasonable performances on micro color difference uniformity and hue linearity[41, 56]. Gamut volume
can also be computed in such a micro-color-difference uniform space. Kunkel proposed an ICtCp-based adaptation hull to evaluate the color difference datasets from Pieri and Pytlarz [56]. Kunkel suggested a method of stacking disks with 1 unit just-noticeable-difference (JND) thickness of a certain adaptation level. Therefore, this is based on stacking volumes from multiple adaptation states, a way to model the best-case sensitivity to dynamic adaptation rather than the steady-state appearance to a fixed adaptation as is done with most implementations of CIELAB and CIECAM02. The study also showed the advantage of the adaptive hull of ICtCp in predicting micro color difference [56]. The good performance from ICtCp and $J_z a_z b_z$ for the small color differences is not surprising given that is what the spaces were designed for. Certainly, within a certain small number of JND steps, the JND’s additivity might be meaningful. However, it is well established that adding JNDs does not lead to a perceptual interval scale for describing large changes in color appearance, which is the goal of color appearance spaces such as CIECAM02. Also usage of JND scales requires some assumptions about adaptation, be it dynamic to the background or static, to make their scales valid. For example, the difference between one gray and (gray+10 JND steps) will likely not to be perceptually equal to that between the gray and (gray-10 JND steps) under a white adaptation background, which is far different from the JND measurement methodology. Therefore, simply adding JND steps straightly up does not produce a meaningful appearance scale. Considering the display presenting most image contents, very of the stimuli will not be within a few JND steps of the instantaneous adaptation level. Moreover, neither ICtCp or $J_z a_z b_z$ includes the chromatic adaptation in the model. Therefore, the color gamut volume expressed in a color appearance model, with reasonable and necessary assumptions about the state of chromatic adaptation, is a more reasonable choice for characterizing the display color reproduction ability.

In addition to the computed 3D color gamut volume in certain color space, the perceptual color gamut volume is necessarily related to the ability to demonstrate colors from the perspective of the normal observers/consumers. Baek published the first study about the
perceptual color gamut volume \[57\]. In Baek’s experiment, observers were asked to evaluate the “rich” (originally in Korean for the observers in South Korean) in two side-by-side images. They examined the relationship between the perceptual color volume (“rich”) and the mathematical 3D color gamut volume in three color spaces: CIELAB, QMh from CIECAM02, and ICtCp. They found the best correlation between the perceptual color volume and the computed color gamut volume in QMh. However, there is some question about the calculations used. For such a type of experiment design, single white point should be used for the entire display, i.e. for both images. Therefore, it is questionable to use two diffuse white levels for the two images presented simultaneously on the same display. The current work has aimed to address this limitation in perceptual evaluation of gamut volume.

This study consists of two parts: metrics for color gamut volume in color appearance models assuming static adaptation, and evaluation of the color gamut volume with two psychophysical experiments. The metrics color gamut volume analysis includes exploration specifications in CIELAB, JCh, and QMh from CIECAM02 and a mathematical model for easy fast computing of the 3D color gamut volumes in these color spaces. The proposed mathematical model predicts the 3D color gamut volume of a display in CIELAB, JCh, and QMh using the peak luminance, RGB primaries of the display, and RGBW configuration parameters. For JCh, and QMh, the model allowed the selection of the viewing environment as dim or dark. The details of these parameters are explained in section 4.2.1. In the second part, two psychophysical experiments were conducted to evaluate the color gamut volume with two different settings accordingly with the two assumptions of diffuse white in the 3D color appearance gamut volume model. The details of the two experiments are introduced after the color appearance gamut volume model. Also included are discussions of the comparison between the experimental results and the color appearance gamut volume model, as well as some important related work.

There has been a long time on the research on displays’ ability in showing colors. Area coverage on chromaticity diagram has been used for a long time in the market for showing the
displays’ color ability. Especially the area coverage on CIE $u'v'$ chromaticity diagram formed by the primaries of the display is used in industry because CIE $u'v'$ is believed to be more uniform compared with CIE $xy$ chromaticity diagram. As explained in the introduction, the chromaticity diagram is just a way to visualize the color boundary while not exactly for any specific quantization application. Mathematically, the chromaticity diagram compressed a 3D data onto a 2D plane. Therefore, it will lose one dimensional data. However, the 3D color gamut is what is used for showing the colors. Therefore, it is more meaningful in exploring the 3D color gamut volume. Both simulations and experiments are proposed to explore the 3D color gamut volume.

4.2 Simulations

Before the experiments, the 3D color gamut volume simulations are proposed. The simulations would focus on the 3D volumes of the displays in specific color spaces and we try to build a model to predict the 3D volume. The color spaces should be in an appearance space, which is more perceptual uniform. There are some absolute color space, more like color coding space, like CIEXYZ, ICIcCp, etc. Two color appearance spaces are used here: CIELAB and CIELAB. Also peak luminance, different color space primaries, and different RGBW configurations are taken into consideration. In the following part, we will introduce simulations parameters, calculations, and result & analysis.

4.2.1 Simulation Parameters

Table 4.1 lists the parameters for simulations. Peak luminance is the maximum white output of the display. Primaries stands the primaries from the color spaces on the chromaticity diagram. Only $xy$ of the primaries from Rec. 709, DCI/P3 and Rec. 2020 are used. Of course, D65 is the white point for all the simulations. Currently most of the HDR displays adopt a RGBW configuration to achieve a higher peak luminance. Therefore, it is necessary
to simulate the impact of different RGBW configurations on the 3D color gamut volume through they have the same peak luminance levels and the same chromaticity diagram area coverage. CLO/WLO represents the RGBW configurations of the display. CLO is short for color light output, which is the white luminance level when the RGB primaries are full on. WLO is short for white light output, which is the white luminance level of the maximum (full RGBW on). If the display/TV is a pure RGB display, CLO/WLO is equal to 1. The lower the CLO/WLO is, the lower of RGB full output over the RGBW. Additionally, diffuse White is an important concept in all color appearance models. Diffuse white is usually used as an anchor point for chromatic adaptation and normalization. Here two different diffuse white settings are used: a constant 200 cd/m\(^2\) and a relative 20% of the peak luminance. 200 cd/m\(^2\) is close to the peak of a traditional standard dynamic range display. Also 200 cd/m\(^2\) is recommended diffuse white level for a 1,000 cd/m\(^2\) peak luminance display by ITU-R [58]. This is also around 20% of the peak luminance. Therefore, a relative 20% of peak luminance is adopted for other peak luminance as well.

### 4.2.2 Calculation Method

Traditional calculation takes sampling in RGB digital count spaces. For example, sampling points in RGB are from [0 0 0] to [1 1 1], with [0 0 0] to [1 0 0], with [0 0 0] to [0 1 0], with [0 0 0] to [0 0 1]. All these ramp data then are converted into specific spaces, CIEXYZ, CIELAB. However, due to the nonlinearity of the color appearance spaces, this would have less sampling points in the dark area. Therefore, a new sampling method is

<table>
<thead>
<tr>
<th>Peak Luminance (cd/m(^2))</th>
<th>200, 300, 500, 1000, 2000, 4000</th>
</tr>
</thead>
<tbody>
<tr>
<td>Primaries</td>
<td>Rec. 709, DCI/P3, Rec. 2020</td>
</tr>
<tr>
<td>CLO/WLO</td>
<td>0.1, 0.3, 0.5, 0.7, 0.9, 1</td>
</tr>
<tr>
<td>Diffuse White Settings</td>
<td>200 cd/m(^2) / 20% of the peak lum</td>
</tr>
</tbody>
</table>

Table 4.1: Parameters for simulations.
used. $L^*$ from CIELAB has a fixed relationship with $Y$. Therefore, the sampling data would be more uniform if the sampling method is on constant even interval $L^*$. Therefore, we sample constant with even interval in $L^*$, which is also sampling in constant $Y$. The key function slicing over the constant lightness/brightness can be found in Appendix A. Moreover, combining this sampling method with gamut ring concept from [59], it will be much easier and quicker to calculate the 3D color gamut volume compared with meshing and calculating the 3D volume in a traditional way.

Masaoka proposed a 2D representation of 3D color gamut volume by slicing the 3D color gamut volume through constant $L^*$ from CIELAB. Accumulating the slices into a 2D area through the constant hue angle would make the 2D area representing 3D volume. The 2D gamut ring concept provides a better visualization of 3D gamut volume. This concepts inspires this new sampling method on the constant $Y$. An example of the new sampling would be demonstrated.

**Example of Simulation**

Following is an example of the new sampling method and a demonstration of the comparison between different RGBW configurations. Figure 4.1 shows the sampling along constant $Y$. The sampling is more condensed in dark end and more sparse in high $Y$ to achieve an even sampling in $L^*$. This is a simulation of DCI/P3 primaries with peak luminance level at 500 cd/m$^2$. Also this is a purified RGB system, where CLO/WLO is equal to 1. Again, in order to have a more even sampling in the nonlinear space, the sampling along $Y$ is more linear in $L^*$ from CIELAB. These $CIEXYZ$ sampling points then are converted into color appearance spaces, like CIELAB, CIECAM02.

Before the conversion, a diffuse white level is required. As stated before, there are two different settings in our simulations: constant 200 cd/m$^2$ and a relative 20% of peak luminance. In this example, they are equal to each other. Before the final demonstration, a resampling procedure will benefit the visual presentation. This procedure resamples the
points along the constant lightness with increment of 1, and along the hue angle from 1° to 360° with increment of 1° as well. Figure 4.2 plots these resampled points in JCh space from CIECAM02.

\[
\text{Vol} = \sum_{L_{\text{min}}}^{L_{\text{max}}} \sum_{h=1}^{h=360} 0.5 \cdot \delta h \cdot (\pi/180) \cdot \text{Chroma}(L, h)^2 \cdot \delta L
\]  

(4.1)

To calculate the volume, simply sum them up according to Equation 4.1, where Cha\(\text{roma}(L^*, h)\) is the chroma as a function of lightness \(L^*(J\) in JCh color space) and hue angle \(h\). \(\delta h\) and \(\delta L^*\) are the increment of hue angle and lightness respectively. In all our simulations, the resampling step would result in \(\delta h = \delta L^* = 1\), which simplifies the calculation as sum of the \(\text{Chroma}^2\) with an coefficient \(0.5 \cdot (\pi/180)\). This would accelerate the calculations by avoiding the traditional meshing and triangulation processing.

In addition to the RGB configuration, different RGBW configurations were also simulated for the practical settings of the commercial available HDR displays. CLO/WLO is a
Figure 4.2: Sampling points in JCh color space from CIECAM02 of DCI/P3 primaries with 200 cd/m² diffuse white setting.

Figure 4.3: Sampling points in JCh color space from CIECAM02 of DCI/P3 primaries with 200 cd/m² diffuse white setting, CLO/WLO = 0.5.
Figure 4.4: Sampling points in JCh color space from CIECAM02 of DCI/P3 primaries with 200 cd/m$^2$ diffuse white setting, CLO/WLO = 0.5.

simple parameter used in the simulations representing the configuration. Here an example of CLO/WLO = 0.5 is demonstrated. Figure 4.3 plots the result of CLO/WLO = 0.5, peak luminance = 500 cd/m$^2$ with 200 cd/m$^2$ diffuse white. Comparison between figure 4.2 and figure 4.3 shows that the volume shape having some concave along the mesh of the gamut, less bright colorful area. The bottom part of the 3D gamut keeps the same.

Moreover, this calculation result can be easily converted into a new 2D visualization of color gamut volume [59]. The method integrates the volume onto the area over the same hue angle. The integration works on constant lightness across all hue angle. These integration points over constant lightness would form a contour, which is called “ring” by the author. Difference between the “rings” is the gamut volume between the two constant lightness. The “ring” concept is similar to the concept of constant-height chart or constant-pressure chart. Figure 4.4 plots the 2D visualization of CLO/WLO = 0.5 on the left and CLO/WLO = 1 on the right. $a^*(RSS)$ and $b^*(RSS)$ are the concepts from [59], which are integrated $a^*$ and $b^*$. The blue contours are the gamut ring for a constant $L^*$, and the red is the ultimated gamut ring of the maximum $L^*$. The 3D gamut volume is shown in the text in the figure. Comparing the two 2D visualization charts can show that they share similarity in shapes but obviously CLO/WLO = 0.5 has a smaller gamut volume and the “ring” stops growing.
when it reaches a point. This can also be found from the two 3D plots, Figure 4.2 and 4.3.

4.2.3 Simulation Result & Analysis

In this section, we will present the simulation results and analysis. Again, all the parameters used in the simulations are listed in Table 4.1. The calculations follow the same procedure in the example above.

Peak Luminance

For both diffuse white settings, there is no significant difference in gamut volume as a function of the peak luminance between JChDark and JChDim or between QMhDark and QMhDim. Therefore, in the following plots, only JChDark and QMhDark are presented. Figure 4.5 shows the 3D color gamut volumes against the peak luminance in three different color spaces for constant 200 cd/m² diffuse white setting. Figure 4.6 plots the same for diffuse white set as 20% of the peak luminance.

For the constant 200 cd/m² diffuse white (Figure 4.5), the 3D color gamut volume increases linearly with peak luminance in CIELCh/CIELAB color space, a slight compression
in JChDark, and a stronger compression in QMhDark. For the CIELCh, the CIELCh formula compresses the absolute luminance level with the power of $1/3$. However the 3D volume should be a third power of the linear scale, which is $(k \times x)^{(1/3)^3} = k \times x$, a linear relationship. For JChDark, the slight different compression parameter in the CIECAM02 model results in a slightly aggressive compression of 3D color gamut volume against the peak luminance. For QMhDark, a stronger compression in the CIECAM02 leads to a stronger compression of the 3D volume.

Figure 4.6: Color gamut volume in CIELCh (left), JChDark (middle), QMhDark (right) against the peak luminance for the diffuse white as 20% of the peak luminance. The solid dots are the simulated data and the dash line is the best fitting 3rd order polynomial function.

For the relative 20% diffuse white setting (Figure 4.6), the 3D color gamut volume in CIELCh/CIELAB is constant in regardless of the peak luminance. It slightly decreases in JChDark with the peak luminance increasing. The 3D color gamut volume in QMhDark showed a compressed increase with the increasing peak luminance. The CIELAB color space is luminance-independent. It normalizes the chosen diffuse white to $Y=100$. The normalization/adaptation in CIELAB would make the color spaces the same in regardless of the absolute luminance level for this relative 20% diffuse white setting. However, for JCh from CIECAM02, a higher absolute luminance level would result in a higher degree of chromatic adaptation level, which normalizes diffuse white. In these simulations, most of
the volume are above the diffuse white level. Therefore, the higher degree of adaptation, the more aggressively these above-diffuse-white were shrunk, then the smaller the 3D color gamut volume is. There is a bit difference from QMh, which adds a luminance-dependent expansion over the JCh, resulting the compressed gamut volume growth as a function of the peak luminance as shown in the plot.

**CLO/WLO**

For different RGBW configurations, peak luminance level, color primary sets, or diffuse white settings did not affect the relationship between normalized 3D volume and the RGBW configuration (CLO/WLO). While the volume against the CLO/WLO shows large variance among the five combinations. Therefore, only the impact of the five different combinations is demonstrated here. Figure 4.7 plots the normalized 3D volume as a function of CLO/WLO. It is an almost even interval between the five 4th order polynomial lines. For JCh and QMh, the dark-condition curve is always above the dim-condition curve. For all the five combinations, the fourth order polynomial fitting well with the simulated discrete points. These fitting polynomials are used in the prediction model.

**Chromaticity Coverage**

Another important settings for a display is the color primaries. The current three popular standard are Rec 709, DCI/P3 and Rec 2020. In all previous simulations, only these three are used. However, 3 would not be enough to derive the relationship between 3D color gamut volume and the chromaticity area. Therefore, a lots of virtual primaries settings are used to explore the relationship between 3D color gamut volume and the chromaticity area coverage. Since $xy$ and $u'v'$ are both used a lot in industrial area for illustrating the color gamut. It is a good chance to explore which has better correlation with the 3D color gamut volume.

In addition to the three standard primaries setting, 1317 virtual primaries settings are used to explore the correlation between 3D color gamut volume and the chromaticity area
Figure 4.7: Normalized 3D color gamut volume in five combinations against CLO/WLO (different RGBW configurations). The solid lines are the best fitting as the 4th order polynomial functions.

Figure 4.8: Virtual primaries sets used for deriving the relationship between gamut volume against chromaticity area coverage.
coverage. Figure 4.8 plots the virtual primaries settings, which are all between Rec. 709 and Rec. 2020. They are evenly selected between the three standards. To simplify the simulations, fixed peak luminance, 4000 cd/m$^2$ and diffuse white, 800 cd/m$^2$, are used. Figure 4.9 plots the normalized color gamut volume against the chromaticity area coverage for both $xy$ and $u'v'$. The text showed the $R^2$. Firstly, it is obvious that the 3D color gamut volume in both CIELAB and JCh has better correlation with area coverage in $xy$ than in $u'v'$. $R^2$ are 0.98 and 0.96 for chromaticity area coverage in $xy$ while they are only 0.47 and 0.52 for the coverage in $u'v'$. For the best fitting-in linear relationship, it should be noted that the offset is not zero. Theoretically, the relationship should be $y = k \times x$ to keep that zero area coverage corresponds with zero color gamut volume. However, this is the best fitting-in between primaries between Rec. 709 and Rec 2020. So this would only be valid for primaries settings between the range. This is more of a practical usage since most of the display can only achieve area coverage between Rec 709 and DCI/P3. In recent years, the simulations between Rec. 709 and Rec. 2020, peak luminance from 500 cd/m$^2$ to 4000 cd/m$^2$, should be enough for usage.
For the argument between $xy$ and $u'v'$, many industrial engineers and researchers claimed that area coverage in $u'v'$ is a better representation of color gamut since CIELUV is more uniform, which is based on the fact that $u'v'$ has better uniformity on MacAdams ellipses. However, MacAdams ellipses are just one measurement data from one single observer on constant luminance level. Color gamut volume in 3D color spaces is much more complex than MacAdams ellipses. Based on our simulations and analysis, the area coverage over $xy$ would be a better representation for 3D color gamut volume than $u'v'$.

4.2.4 Mathematical Model & Verification

Mathematical Model

The previous several sections presented the relationship of the 3D color appearance gamut volume in five different combinations against the peak luminance, the primary set ($xy$ chromaticity area coverage), and the RGBW configuration (CLO/WLO) respectively. The mathematical model in Eq. 4.2 is simply a multiplication of three ratios with a constant. The $PeakLumRatio$, $xyRatio$, and $RGBWRatio$ are the three ratios as a function of peak luminance, $xy$ chromaticity area coverage, and CLO/WLO respectively. $Rec2020_{constant}$ is a constant of 3D color appearance gamut volume of 1000 cd/m$^2$ peak luminance, 200 cd/m$^2$ diffuse white setting (also 20% of the peak luminance), RGBW configuration with CLO/WLO of 1, and Rec. 2020 primary set. For the three ratios, polynomial functions are used to calculate the relationships. The detail of the polynomial functions are provided below. Before showing the calculations for the three variables, the constant $Rec2020_{constant}$ for the five color space combinations are listed in Table 4.2.

$$Vol_{pred} = PeakLumRatio \times xyRatio \times RGBWRatio \times Rec2020_{constant} \quad (4.2)$$

A third order polynomial function was used for deriving the $PeakLumRatio$. Table 4.3 lists the parameter $P1$, $P2$, $P3$, $P4$ for calculating $PeakLumRatio$ using Eq. 4.3, where $x_p$
<table>
<thead>
<tr>
<th>Color appearance space</th>
<th>Rec2020 constant</th>
</tr>
</thead>
<tbody>
<tr>
<td>CIELCh/LAB</td>
<td>0.97587 * 10^7</td>
</tr>
<tr>
<td>JChDark</td>
<td>0.5534 * 10^7</td>
</tr>
<tr>
<td>JChDim</td>
<td>0.7547 * 10^7</td>
</tr>
<tr>
<td>QMhDark</td>
<td>0.6633 * 10^7</td>
</tr>
<tr>
<td>QMhDim</td>
<td>0.7606 * 10^7</td>
</tr>
</tbody>
</table>

Table 4.2: Values of Rec2020 constant for the five combinations.

<table>
<thead>
<tr>
<th>Constant diffuse white</th>
<th>CIELCh</th>
<th>JChDark</th>
<th>JChDim</th>
<th>QMhDark</th>
<th>QMhDim</th>
</tr>
</thead>
<tbody>
<tr>
<td>P1</td>
<td>0</td>
<td>0.0169</td>
<td>0.0083</td>
<td>0.0286</td>
<td>0.0244</td>
</tr>
<tr>
<td>P2</td>
<td>0</td>
<td>-0.1558</td>
<td>-0.0860</td>
<td>-0.2423</td>
<td>-0.2131</td>
</tr>
<tr>
<td>P3</td>
<td>1.0097</td>
<td>1.0836</td>
<td>1.0613</td>
<td>1.0295</td>
<td>1.0556</td>
</tr>
<tr>
<td>P4</td>
<td>-0.0122</td>
<td>0.0589</td>
<td>0.0179</td>
<td>0.1911</td>
<td>0.1385</td>
</tr>
</tbody>
</table>

Table 4.3: Polynomial parameters for calculating PeakLumRatio.
is the peak luminance in units of cd/m$^2$. For CIELCh, it is a linear relationship for constant diffuse white and a constant value for a relative 20% diffuse white.

$$PeakLumRatio = P1 \times (x_p/1000)^3 + P2 \times (x_p/1000)^2 + P3 \times (x_p/1000) + P4$$  \hspace{1cm} (4.3)

A linear model was used to compute the $xyRatio$ as shown in Eq. 4.4, where $x_{xy}$ is the area coverage of the triangle connecting the three primary on $xy$ chromaticity diagram. The constants are provided in Table 4.4. For the five combinations, we found the linear relationships are very close to each other among JChDark, JChDim, QMhDark, and QMhDim. Therefore, the same parameters are used for the four.

$$xyRatio = P1 \times x_{xy} + P2$$  \hspace{1cm} (4.4)

<table>
<thead>
<tr>
<th></th>
<th>CIELCh</th>
<th>JChDark, JChDim, QMhDark, QMhDim</th>
</tr>
</thead>
<tbody>
<tr>
<td>$P1$</td>
<td>5.2540</td>
<td>4.3988</td>
</tr>
<tr>
<td>$P2$</td>
<td>-0.1547</td>
<td>0.0037</td>
</tr>
</tbody>
</table>

Table 4.4: Polynomial parameters for calculating $xyRatio$.

For $RGBWRatio$ a fourth order polynomial is used for calculation. Eq. 4.5 and Table 4.5 show the equation and parameters respectively. The $x_c$ is the CLO/WLO used to characterize RGBW display. Again, the CLO/WLO is a variable used to characterize the RGBW configuration. The measurement of the CLO/WLO should be depending on the actual display’s manufacture. It is assumed to be a measurable value for each display. However, it may vary with different mode, i.e., power saving mode, safe mode etc, which is not in the scope of this paper.

$$RGBWRatio = P1 \times x_c^4 + P2 \times x_c^3 + P3 \times x_c^2 + P4 \times x_c + P5$$  \hspace{1cm} (4.5)
<table>
<thead>
<tr>
<th></th>
<th>CIELCh</th>
<th>JChDark</th>
<th>JChDim</th>
<th>QMhDark</th>
<th>QMhDim</th>
</tr>
</thead>
<tbody>
<tr>
<td>P1</td>
<td>-0.5067</td>
<td>-0.6718</td>
<td>-0.5707</td>
<td>-0.8737</td>
<td>-0.7990</td>
</tr>
<tr>
<td>P2</td>
<td>0.6454</td>
<td>1.4451</td>
<td>1.0617</td>
<td>2.1797</td>
<td>1.9066</td>
</tr>
<tr>
<td>P3</td>
<td>-0.7038</td>
<td>-1.7809</td>
<td>-1.3283</td>
<td>-2.5875</td>
<td>-2.2926</td>
</tr>
<tr>
<td>P4</td>
<td>1.5786</td>
<td>1.9803</td>
<td>1.8324</td>
<td>2.1694</td>
<td>2.1107</td>
</tr>
<tr>
<td>P5</td>
<td>-0.0132</td>
<td>0.0274</td>
<td>0.0051</td>
<td>0.1120</td>
<td>0.0743</td>
</tr>
</tbody>
</table>

Table 4.5: Polynomial parameters for calculating $RGBWRatio$.

3D color gamut volume remains difficult and time-consuming to compute accurately and precisely. This model provides a simple method to compute the 3D color gamut volume in all the 5 combinations.

**Verification**

In order to evaluate the model’s performance, a separate set of 1000-simulated-displays verification data was created. The 1000-display verification data have random peak luminances between 200 and 4000, random CLO/WLO between 0.1 and 1, random primaries set ($xy$ chromaticity area coverage) between Rec. 709 and Rec. 2020. 3D color gamut volume for both the constant 200 cd/m$^2$ and the relative 20% diffuse white settings were computed. Therefore, in total there are 2000 independent gamut volumes for verification. Figure 4.10 plots the predicted 3D color gamut volume from the model (Eq. 4.2) against the ground truth 3D color gamut volume from calculations. The dash line is $y = x$ perfect performance line. All data were along the perfect $y = x$ line. Table 4.6 lists the mean of the absolute error ($|Vol_{pred} - Vol_{truth}|/Vol_{truth} \times 100\%$) and the standard deviation of the absolute error for all the five combinations. The mean error is not more than 2.4% with no more than 1.8% standard deviation. These statistical data validate the accuracy of the model (2.4% ± 1.8%).
<table>
<thead>
<tr>
<th></th>
<th>CIELCh</th>
<th>JChDark</th>
<th>JChDim</th>
<th>QMhDark</th>
<th>QMhDim</th>
</tr>
</thead>
<tbody>
<tr>
<td>mean</td>
<td>1.73%</td>
<td>1.83%</td>
<td>1.73%</td>
<td>2.40%</td>
<td>2.36%</td>
</tr>
<tr>
<td>STD</td>
<td>1.33%</td>
<td>1.42%</td>
<td>1.30%</td>
<td>1.78%</td>
<td>1.77%</td>
</tr>
</tbody>
</table>

Table 4.6: Mean errors of the verification simulations.

Figure 4.10: True color gamut volume against the predicted volumes for random 200 virtual primaries settings of random peak luminances with $CLO/WLO$ as 0.1, 0.3, 0.5, 0.7, 0.9 and 1.
4.3 Experiments

In addition to the simulations and the mathematical model above, psychophysical experiments are necessary to explore the perceptual color gamut volume, and compare that with the theoretical simulation results. As stated before, our studies focus on the dynamic range displays. Moreover, as with the simulations, the black level would not change the color gamut volume too much since they are very small in terms of volume, even though this is a big difference for dynamic range. Therefore, the peak luminance of the display would affect the color gamut volume more. Our experiments would aim at the perceptual color gamut volume when displays have different peak luminance levels. Also as with the simulations above, two different diffuse white levels should be used in the experiments: a constant diffuse white level and a peak-relative diffuse white level. The proposed experiments are introduced in the following sections.

4.3.1 Experiment I Design

In this experiment, Figure 4.11 illustrates the 9 images for experiments: including 3 natural HDR images, and 6 calibrated HDR images. The 3 natural images include typical high dynamic range images: dark scene with bright fire, bright flowers, and super bright sun with the beach. The 6 images are calibrated to 200 peak luminance diffuse white using a color checker when shown on a 1000 cd/m$^2$ peak luminance display. The display used for the experiment is a professional 30 inch SONY OLED trimaster, which can achieve a very stable 1000 cd/m$^2$ peak luminance. For each original image, the peak luminance is clipped from 1000 cd/m$^2$ to 500 cd/m$^2$, 300 cd/m$^2$ and 200 cd/m$^2$. The clipping is based on clipping the corresponding RGB channel while not just the total luminance channel. This is trying to simulate the displays’ ability showing the same diffuse white level but different abilities in showing bright colors/highlights above diffuse white.

The method used for this experiment is paired-comparison. For each original image,
Figure 4.11: 9 images for experiments, image 1 - 3 are natural images and image 4 - 9 are images with calibrated diffuse white level.

4 images would result in 6 pairs, in total $6 \times 9 = 54$ pairs. Each observer would make all 54 judgements over the pairs. For pair-comparison, the first image would show for 15 seconds, followed by a 10 sec mid-gray, after which the second image will show for another 15 seconds. The sequence in each pair was randomized. Such a design was to simulate the perspective of a customer, who makes judgement between two displays in different areas, such as one at home and the other at a friend’s. The reason that we did not use the side-by-side pair-comparison is that it is well known that observers can adapt to their own devices. The side-by-side pair-comparison would exaggerate the effect from peak luminance in the perspective of a customer. Observers were supposed to pick the one more colorful and more detailed. Theoretically, the display with larger color gamut volume should display the image more colorful and more details.
4.3.2 Experiment II Design

The second experiment focused on the perceptual color gamut volume of different peak luminance with a relative diffuse white level. The same 9 images (Figure 4.11) were used in this experiment. For each original image, the whole image was filtered by ND filters of transmittance as 0.7, 0.5 and 0.35. Each original image would result in 4 images as well. The 4 images would generate 6 pairs. Therefore, there are 54 pairs. This design is trying to simulate displays with similar configurations but different peak luminances of all the primaries. Therefore, different configurations would scale the whole image. Again, the sequence within a pair is: 15 seconds of the first image, 10 second the midgray, 15 seconds of the second image. The sequence of each pair was randomized. Also the order of the 54 pairs are also randomized. The same question was used here for observers, picking the one more colorful and more detailed.

4.3.3 Experiment Result Analysis

In this result and analysis section, the results of the two experiments will be demonstrated separately. Following that, the comparison between the two experiments will be discussed.

Experiment I

Figure 4.12 plots the perceptual color gamut volume against linear peak luminance on the left and that against log scale of peak luminance on the right. The error bar is the 95% confidential interval based on Montag’s work[60]. Firstly, pair-comparison gives out the interval scale result. As introduced in the background chapter, the interval scale is only meaningful in terms of the difference. Any absolute value is not meaningful, especially zero. Here z-score would be meaningful as perceptual color gamut volume, however not in a numerical way. From the plot, it is easy to find that the perceptual color gamut volume, “colorfulness and detailed”, is more linear to log scale of the peak luminance. This
is the average result of the 9 images.

Figure 4.12: Perceptual color gamut volume against linear peak luminance (left) and log peak luminance (right) for experiment I.

Figure 4.13 plots the result for individual image as z-score against log scale of peak luminance. Here z-score is the same mean to the perceptual color gamut volume. However, the perceptual color gamut volume is more meaningful for the display. Therefore, for each image, z-score is used for individual image. In each subplot, the thumbnail image illustrates the luminance distribution of the original image when shown on the display. The grayscale indicates the luminance level under 200 nit; the red indicates the luminance between 200 cd/m² and 300 cd/m²; the green means the luminance between 300 cd/m² and 500 cd/m²; and the blue areas are those of luminance level between 500 cd/m² and 1000 cd/m². These thumbnail images showed directly where the clipping happens. For example, the sun is clearly clipped when the peak luminance is clipped from 1000 cd/m² to 500 cd/m² (blue area will lose any texture or detail. The details in the blue and the green areas will be washed out when the peak luminance is clipped to 300 cd/m². All colored area will lose texture when the peak luminance is clipped to 200 cd/m². So the lower the peak luminance is, the more
Figure 4.13: Perceptual color gamut volume (z-score) against log peak luminance for experiment I.
clipping it will be. The surrounding of the sun will lose texture gradually with aggressive clipping. Image #6 is another good example. The texture of the sweater is bleached when the peak luminance is clipped to 300 cd/m² and 200 cd/m² but not for 500 cd/m². Similar can be found on the sweater in image #8. Image #1 is a low key image. The fire will lose detail gradually when the peak luminance decreases.

For the individual result, z-score showed no difference between 200 cd/m² and 300 cd/m² for image #1. The thumbnail also showed that very little pixels having luminance level between 200 cd/m² and 300 cd/m² (red). Another interesting example is image #6. The luminance level distribution showed that a large area are between 200 cd/m² and 300 cd/m². A big difference of z-score between 200 cd/m² and 300 cd/m² showed agreement with the luminance segmentations. Image #9 demonstrates another type of image. Firstly image #9 is a complex scene, with some trees in front and buildings on the back. There is slight sparse areas having luminance level above 200 cd/m² on the back buildings. Moreover, these areas are very close to uniform gray. Therefore, the clippings would not affect these too much. This slight difference in the back is not easy to be captured by observers especially under this one by one test methodology. Therefore, there is almost no difference between the four different peak luminance levels. So, the flat z-score of the four different peak luminance is not a surprising result. It means that there is no difference in the perspective of observers even some details/texture on the small area on the back of the image was bleached by the clipping.

Experiment II

Figure 4.14 plots the perceptual color gamut volume against the linear peak luminance level (left) and the average against the log scale of peak luminance (right) for the experiment II. Similar to experiment I, the average result against the log scale of peak luminance levels presented better linearity compared to the average result against the linear of peak luminance levels.
Figure 4.14: Perceptual color gamut volume against linear peak luminance (left) and log peak luminance (right) for experiment II.

Figure 4.15 showed the z-score against the log scale of peak luminance levels of each image. Overall the result of each image showed very good linearity against log scale of the peak luminance levels. But the results of image #3, #6, and #9 showed a clear nonlinearity. Image #3, and #6 showed a curve as a power function larger than 1. Result of image #9 is closer to a compression function. For both image #3 and #6, there are a large areas of high bright area as the key area in the images, the sun for image #3 and the white skirt for image #6. For low key image with bright area, this area became observers’ attention. This was supported by the observers that they claimed focusing on the sun for image #3 and the white skirt for image #6. The focus area is very important for this measurement methodology. One-by-one methodology requires observers’ memory of the first image. The time-limit enhanced the importance of focusing area. When ND filter was applied, a stronger difference between the pair was perceived by observers. However, there is only very small and spacious area in image #9 includes high bright area. Image #2 also has a similar luminance distribution, spacious bright area. Also image #2 is a complex high key image.
Figure 4.15: Perceptual color gamut volume (z-score) against log peak luminance for experiment II.
In this experiment, the neutral density filters were used. As is known, human vision system adapts to the absolute luminance level. But colorfulness is a metric that is associated with absolute luminance level. Colorfulness increasing with luminance level is known as Hunt effect [61]. The average results in both experiments verify Hunt effect. Perceptual color gamut volume ("colorfulness and detailed") increases with absolute luminance level. For the specific image, the luminance distribution of the image may have an impact on the exact relationship. A large bright area may strengthen observers’ perception of the image. When the ND filters were applied, a stronger difference between the pair of images was perceived. On the other side, a complex and relative luminance uniform image would cause a weaker effect. Therefore, this could result in such a difference. This explanation is supported by the feedback from the observers. Most of the observers reported that they could not see much difference for image #9 for most cases, but they could feel strong difference when the ND filters were applied as the sun was so bright. This is a higher level of Stevens’ law, that the cognitive function of observers’ attention has impact on the strength of Stevens’ law.

4.3.4 Experiment Discussion

Firstly, in both experiments, the perceptual color gamut volume is interpreted as “colorfulness and detailed”. In psychophysical experiments, it is very important to describe the terminology for observers. We believed that a display with larger color gamut volume should have a fundamental ability in reproducing the same image data more colorful and in details. It should be noted that the same image data without any further processing should be used as is known some HDR tone mapping algorithms can be used to make an image more detailed. However, in our experiments, we focused on the fundamental ability of the displays. Therefore, only hard clipping is used to process the original image data when the luminance level of the original image is beyond the display’s ability.

Direct analysis data from the psychophysical experiment is z-score, which is an interval scale based on the pair-comparison methodology. Z-score can also be called the perceptual
color gamut volume. In both experiments, the perceptual color gamut volume showed a better linearity against the log scale of peak luminances compared with the linear peak luminances. It is reasonable considering the content of the images used in experiments. In a pictorial images, most area of the image is under diffuse white as matte material. Therefore, in experiment I, the clipping (different peak luminance levels) does not impact most area of the image but only the highlight areas. For most areas above the diffuse white, there are three different types: 1) highlights. For example, the highlights from the instrument in image #5 in Figure 4.11. 2) direct lights. For example, fire in image #1, sun in image #3 from Figure 4.11. 3) the area is under the bright area due to the nonuniformity of illuminating in the scene. For example, the white sweater in image #6 from Figure 4.11. Almost all of the three situations have only a small area in an image. These are the areas that would affect observers’ choice. Therefore, the perceptual brightness/colorfulness of these area should have a better linearity with the perceptual color gamut volume. As is known, perceptual brightness is closer to a power function of absolute luminance level. Therefore, perceptual brightness of the peak luminance level would have a better linearity with the log scale than the linear of the peak luminance. In experiment II, the ND filter scaled the whole images. There are two explanations. One is the overall brightness of the image, and the other is the peak brightness has a significant impact. More analysis of the average image brightness and peak brightness should be done. This could be a good explanation for the difference between the two experiments.

In addition to the peak luminance and log10 of the peak luminance, several more parameters were added for analysis, including 3D color gamut volume in LCh, JChDark, QMhDark, peak Q and the log10 scale of these. According the image processing of the two experiments, for experiment 1 (peak luminance hard clipping), a constant diffuse white level was adopted, and for experiment 2 (filters), a relative 20% diffuse white was used. A study about the perceptual diffuse white estimation supports these diffuse white settings [62]. $R^2$ values between the perceptual color gamut volume and these parameters are listed in Table 4.7. The bold
Table 4.7: $R^2$ between the parameters and the perceptual color gamut volume.

<table>
<thead>
<tr>
<th></th>
<th>Exp1</th>
<th>Exp2</th>
<th>average</th>
</tr>
</thead>
<tbody>
<tr>
<td>Peak Luminance</td>
<td>0.862</td>
<td>0.974</td>
<td>0.968</td>
</tr>
<tr>
<td>log10(Peak Luminance)</td>
<td>0.974</td>
<td>1.000</td>
<td>0.987</td>
</tr>
<tr>
<td>LCh</td>
<td>0.861</td>
<td>0.000</td>
<td>0.431</td>
</tr>
<tr>
<td>log10(LCh)</td>
<td>0.975</td>
<td>0.000</td>
<td>0.788</td>
</tr>
<tr>
<td>JChDark</td>
<td>0.889</td>
<td>0.956</td>
<td>0.923</td>
</tr>
<tr>
<td>log10(JChDark)</td>
<td>0.976</td>
<td>0.956</td>
<td>0.966</td>
</tr>
<tr>
<td>QMhDark</td>
<td>0.914</td>
<td>0.989</td>
<td>0.952</td>
</tr>
<tr>
<td>log10(QMhDark)</td>
<td>0.975</td>
<td>0.999</td>
<td>0.987</td>
</tr>
<tr>
<td>Peak Q</td>
<td>0.957</td>
<td>0.992</td>
<td>0.973</td>
</tr>
<tr>
<td>log10(Peak Q)</td>
<td>0.974</td>
<td>0.998</td>
<td>0.986</td>
</tr>
</tbody>
</table>

Figure 4.16: Normalized parameters against the perceptual color gamut volume for experiment 1 (left) and experiment 2 (right).
numbers indicate the best three parameters, $\log_{10}(\text{Peak Luminance})$, $\log_{10}(\text{QMhDark})$, and $\log_{10}(\text{Peak Q})$. The colorfulness of an image would increase as the absolute luminance level, the well-known Hunt effect [63]. LCh and JChDark do not predict colorfulness. Hence, LCh and JChDark did not demonstrate a good performance as expected. QMh takes the absolute luminance into account, and the $M$ is the colorfulness attribute from the CIECAM02 model. It is not surprising that QMh performs better than the LCh and JCh. Figure 4.16 plots the best 3 parameters against the perceptual color gamut volume for both experiments. Firstly, the linearity is much better in experiment 2 than in experiment 1, the same as demonstrated by the $R^2$. For experiment 1, it can be found that the difference between the four different peak luminance levels is only around 5% for the $\log_{10}(\text{QMhDark})$ and $\log_{10}(\text{peakQ})$, but 25% for $\log_{10}(\text{peakLum})$. Very similar variance data of the normalized parameters in experiment 2. Another interesting result about the plots is the line of $\log_{10}(\text{peakQ})$ is above that of $\log_{10}(\text{QMhDark})$ in experiment 1 while the order is reversed for experiment 2. However, since the two experiments were conducted separately it is difficult to have a direct comparison of the two experiments. Therefore, the order of the lines is not that meaningful based on the current available experimental data.

There was very few experiments directly addressing the perceptual color gamut volume. Baek and Kwak conducted a psychophysical experiment evaluation of the 3D color gamut volume using real images [57]. The experiment used the word “rich” in color. The real word used in experiment was a Korean word as all the observers were Koreans. The “rich” in color is similar with our “colorfulness”. However, the biggest difference between their experiments and our experiments is the measuring methodology that Baek’s experiment adopted a side-by-side trial but our was one-by-one. The experiment methodology reflects the situations that the result could be applied. As stated before, our goal is such as to the situation that observers have visual experiences of different displays spatially separated, i.e. allowing adapting to different displays. Baek’s experiment would reflect more of observers’ visual experiences when displays were side by side, like in retail store. Baek reported a
better linearity between 3D color gamut volume in $Q_{a_m}b_m$ and the perceptual color gamut volume. However, the calculation about the theoretical 3D gamut volume is questionable. The side-by-side methodology would put the two images under the same viewing conditions, thus the same adaptation conditions. The same diffuse white should be used in calculations. However, peak luminance of each image was used for diffuse white in the calculations. They also claimed a similar good correlation between the perceptual color gamut volume and the 3D color gamut volume in ICtCp space. The ICtCp color space is not an appearance model, more of an encoding space. We may include that into the computation as future work.

4.4 Summary

In this chapter, a mathematical model for 3D color gamut volume in color appearance spaces, CIELAB and JCh, QMh from CIECAM02 was proposed. A new sampling method was used, which is based on the new 2D visualization of the 3D color gamut volume [59]. This new sampling method makes the sampling more uniform across in CIELAB space, which would make the sampling and interpolation more efficient. Also this would be easily converted to the new 2D visualization. The 2D visualization proposed by Masaoka would illustrate a 3D color gamut in 2D diagram, which has the potential to be a useful tool for display engineers and technicians.

The analysis of a series of simulations demonstrated some interesting results. Firstly, the 3D color gamut volume in CIELAB has a linear relationship to the peak luminance in regardless of the diffuse white level. There is very slight concave relationship between the color gamut volume in JCh and the peak luminance for a constant diffuse white, and strictly linear relationship between them for a relative 20% diffuse white. In most current HDR WCG displays, they have a white optical channel in addition to the traditional RGB channels to achieve a higher peak luminance level. $CLO/WLO$ is a parameter used to incorporate such a configuration. The analysis of $CLO/WLO$ from 0.1 to 1 showed a stable relationship
regardless of the peak luminance level of color primaries. This suggested that CLO/WLO is an independent parameter in the 3D color gamut volume. Certainly, there is slight difference between that in CIELAB and in JCh, and QMh. The simulations demonstrated a surprising result about the relationship between the 3D color gamut volume and the chromaticity coverage. Simulations of 1317 virtual primaries showed a surprisingly good correlation ($R^2 > 0.95$) between the 3D color gamut volume and the $xy$ chromaticity area coverage, and very poor correlation ($R^2 \approx 0.5$) between the 3D color gamut volume and the $u'v'$ chromaticity area coverage. As is known, MacAdam’s ellipse [45, 64] has a better uniformity in $u'v'$ chromaticity than $x''y$, which has been used as a support for claiming $u'v'$ chromaticity diagram has better uniformity than $xy$ chromaticity diagram. However, MacAdams’s data is only one person with constant luminance/lightness. Considering the 3D color gamut volume as a whole, it is not surprising that $u'v'$ does not perform as well as in MacAdams’s ellipses. This is also important for industrial engineers and technicians as they have been using $u'v'$ in indicating color difference or color gamut. In the mathematical model, the predicted 3D color gamut volume is the production of $xyRatio$, $peakLumRatio$, $rgbwRatio$ with $Rec2020Vol$. An independent dataset was used to validate the model. The independent verification data showed a 3% error of the mathematical model. This error makes the model a practical usage tool for an easy, fast calculation of 3D color gamut volume.

In addition to the theoretical calculation of 3D color gamut volume, two psychophysical experiments were designed and conducted to explore the effect of peak luminance over the perceptual color gamut volume. Experiment I explored the perceptual color gamut volume of a peak luminance clipping over 9 HDR images. The clipping would only affect the luminance channel while keep the rest intact. Observers were asked to make judgement based on “colorfulness and detailed”. The average of the result showed a good linearity against the log scale of peak luminance. According to the feedback of the observers, they reported more over the “detailed”, the loss of details of the bright area. In experiment II, ND filters were used to scale the whole image. The average result also showed a good linearity against log scale
of the peak luminance. In this experiment, observers tended to make decision more over “colorfulness” as they reported that the overall brightness changed resulting in the change in colorfulness but not much loss of details. Also comparing the two results of the two experiments presented that a steeper slop linearity in experiment II than that in experiment I. This is showing the possible of the overall brightness effect over the result. This could be a future work.

Figure 4.17: Illustration of used volume by a pictorial image. The two black lines present the two opponent channel in most color appearance models, and the vertical axis (Z-axis) stands for the achromatic channel lightness/brightness.

The two experiments simulated two different image rendering methods, clipping and scaling. They are also corresponding with the two diffuse white settings in the model. The result of the experiment is not consistent with the simulations. The 3D color appearance gamut volume model predicts a linear relationship between the 3D volume and the peak luminance. While the experimental result showed better linearity between the perceptual color gamut volume and the log(peak luminance). This is not a surprising result considering the difference between the model and the experimental methodology assessing the color
volume. The perceptual color gamut volume can only be measured through experimental images on the displays. However, the pictorial images would not use the display’s whole 3D gamut entirely. Actually, most of the images would be of reflective objects with some highlights, direct lights, the reflective objects close to lighting source, etc. Therefore, most of the content/pixels in a pictorial image are under a certain diffuse white level. Figure 4.17 illustrates the concept, where the gray lines represent two opponent color channels as in almost all color appearance spaces and the vertical axis (Z-axis) shows lightness/brightness. The colored 3D ellipsoid represents the volume used for reflective object and the three spikes above that are volume for highlights, etc. For experiment I, the hard clipping would mostly affect these spikes. Therefore, observers’ response should be more similar to the impact on these spikes. The log scale of the peak luminance is obviously a better metric, than the linear peak luminance, to describe the observers’ response, which is more linear to the log scale of luminance. It should be noted that this is based on a hard clipping image rendering. Utilizing a tone-mapping-operator (TMO)/soft clipping, which was designed to map a higher dynamic range image into a lower dynamic range display while preserving the details as much as possible, would probably be different from this experiment. A TMO would optimize the usage of a limited color gamut volume. A TMO is a mapping method of nonlinear compression from the dynamic range to the display’s dynamic range while aiming at preserving more contrast information. An easy way to understand the TMO is that one TMO applies a slight compression over the 3D colored ellipsoid and a higher compression over the spikes over the 3D ellipsoid in Figure 4.17. The compression can be applied globally or locally according to the specific TMO algorithm. However, that is out of the scope of this paper but should be taken into consideration for any specific display is using TMOs in processing and display of high dynamic range image/video content. For experiment II, the ND filter was simulating the scaling processing. The filters reduce the whole image in the same scale. QMh from CIECAM02 model was designed to predict the well known Hunt effect [63]. It is not surprising that the 3D volume in QMh works better than LCh and JCh.
The only known study about evaluating the color gamut volume was reported by Baek [57]. However, as mentioned in the introduction, the calculation of the volume is questionable considering the chosen white levels. Therefore, we will not cross compare the results directly. However, the judging criteria/question for the observers in the psychophysical experiments is worth discussion. The question posed to the observers is critical for a psychophysical experiment. For example, there is a clear difference between color fidelity and observers’ color preference in image. In these two experiments, a larger perceptual color gamut volume is supposed to demonstrate the image as more colorful and more detailed with the same image source data. The images were encoded in 10 bit HLG, which should be free of any artifacts as in the pilot test. This is also the main reason that ND filters were used instead of scaling the images digitally. Certainly, there is no standard definition of perceptual color gamut volume, and it could vary from application to application. For example, Baek’s work using ‘richness’ in Korean word during their experiment [57]. Therefore, some variance would be expected for different questions. Moreover, for each image, 6 pairs were evaluated by all observers as a balance between the fatigue of observers and the efficiency. For $R^2$ analysis of the four points, the difference between different metrics in Table 4.7 is not that big. More pairs can be evaluated in the following study. Also, the two experiments are only assessing the displays with different peak luminance levels. As simulated in the model, the primary set has a great impact on the 3D volume. It is necessary to explore the effect from the primary set and the joint effect of primary set and peak luminance. While difficult to implement and test, future studies of dynamic gamut computations considering spatially- and temporally-local adaptation would also give insight into more aspects of real-world display performance.
Chapter 5

Perceptual Diffuse White in HDR images when on HDR display

5.1 Introduction

It is known that chromatic adaptation is an important mechanism of human visual perception. In most current color appearance models, from the simplest CIELAB to CIECAM02 [48], or any other color appearance models [46], chromatic adaptation transforms require one to adopt the "white" as an adapting point, or anchor. The "white" here refers to diffuse white / matte white, which is the white level from a perfect diffuser. Usually for the reflective objects, the uniform neutral matte object with highest possible reflectivity is considered as diffuse white in all current color appearance models. It should be noted that the diffuse white/matte white should be separated from the highlight white, which is a highly reflective specular object and often perceived as a light source or self luminous. Moreover, "white" here stands for the white level of an image when presented in the display, while not the actual white of the object. The “white” here is all referring to the reproduced white in scene.

Often the peak luminance of the display is used as diffuse white in display characterization. In the classic Berns’ and Day’ display colorimetric methods [65–67], the peak of the
display was used as diffuse white in the colorimetric color difference error minimization. It is a reasonable assumption as in standard dynamic range (SDR) display peak luminance is normally close to the diffuse white when presenting real images. For many current HDR displays, a special "white" optical channel is added to achieve a higher peak luminance. The peak luminance is usually used to demonstrate the highlights, light sources, or bright high-chroma colors. For these HDR displays, it would not be a reasonable assumption that the peak luminance is the diffuse white in characterizing the display or image appearance. Moreover, in evaluating 3D color gamut volume of the display, the diffuse white is important for all color appearance models. Therefore, determining the diffuse white is a fundamental factor in calculating the 3D color gamut volume. In Masaoka’s initial research about 3D color gamut volume of HDR displays, different luminance levels were used as "white" in the simulations [52]. In Jiang, et al.’s following research about building a mathematical model predicting 3D color gamut volume of HDR wide-color-gamut (WCG) displays, two assumptions about "white" level were made, constant 200 nits and 20% of the peak luminance [68]. Additionally, ITU-R made a recommendation for using around 20% of peak luminance in HLG encoding system for 1000 nits peak luminance HDR WCG displays [58]. Additionally, Baek published the first study about the perceptual color gamut volume [57]. In Baek’s experiment, two different "white" levels were used in calculations and compared with experiment data for side-by-side image evaluation. This means two different "white"s were used in a single frame on one display, which is questionable. Jiang, et al. presented another study about the perceptual color gamut volume through a psychophysical experiment [69]. In Jiang’s experiment, the paired images were presented one by one separately with a few seconds interval. Theoretically, different diffuse white levels could be used for the two images. However, the perceptual estimation of diffuse white did not show too much difference between the pair of the same image content and different peak luminance levels [69].

Three experiments are included in this study. The first and the second experiment focused on the perceptual evaluation of diffuse white level and the third experiment showed
the impact of measuring methodologies on the absolute luminance level of the estimated
diffuse white when HDR images were presented on displays. Experiment I was partially
reported in [69]. It should be noted that our experiments are different from the research
about lightness perception in HDR images. Allred’s research focused on mapping HDR to a
series of lightness, where the highest patch was always perceived as white [70]. This study
explores diffuse white in HDR images when presented in displays and comparing that with
the calibrated diffuse white in HDR images.

5.2 Experiment Design

Experiments I and II adopted the same configuration as shown in Figure 5.1, a SONY 30
inch Trimaster 4K display with two black plates blocking any ambient lights. This SONY
display could reach a stable 1000 nits peak luminance. However, the SONY display could
not reach full screen 1000 nits due to the Average Picture Level (APL) limitation. In our
experiments, the APLs of the images are low enough that the display could reach a stable
1000 nits peak luminance. A LG 55-inch OLED TV was used in experiment III, Figure 5.4.
In all experiments, the observers were sitting in a dark room with curtains or stands blocking
the flare.

5.2.1 Experiment I

Images for Experiment

Figure 5.2 shows the images used in experiment I. All 6 images were taken with a well-
calibrated HDR camera. During capture, a color checker was used to calibrate the exposure.
The exposure was following the ITU-R suggestion that the diffuse white should be calibrated
to 200 nits for 1000 peak luminance. All 6 original images would have a 1000 nits peak
luminance when shown on the SONY display. The peak luminance of these 6 images was
clipped to the additional three different experimental levels: 500 nits, 300 nits, and 200 nits.
Figure 5.1: Configuration for experiment I and II, two black matte stands were used to block any possible flare on the display.

Therefore, there were 24 images for evaluation. A five-level methodology was used in this experiment. In each image five gray patches were added at the center. Figure 5.3 illustrates an example of the five levels. The five level patches were padded with black, with width of 20% of the patch size. The absolute luminances of the five levels were: 176 nits, 285 nits, 439 nits, 640 nits, and 879 nits.

Observers

In total, 21 observers participated in the experiment, including 10 expert observers and 11 naive observers. The expert observers were identified by their background knowledge of a perfect diffuser/lambertian surface in radiometry. They were graduate students majoring in imaging science or color science, and faculty members in related area.
Observer’s Task

Each observer was asked to use a keyboard indicating their estimation of the diffuse white. They were asked to pick two consecutive levels, where their estimation lay in between (or pick the same level twice if they believe their estimations were very close this level). All observers were instructed regarding the difference between diffuse white and specular white/highlight white before the experiment.

5.2.2 Experiment II

Images for Experiment II

The same 6 original images, Figure 5.2, were used in experiment II. For each original image, three different neutral density filters were used for scaling the whole image luminance level. The transmittances of the three filters are: 0.7, 0.5, and 0.35. Including the original one, each image generated 4 images, in total 24 images for observation. As discussed before, these images were calibrated with 200 nits diffuse white level when shown on the display. Therefore, for different filters, the calibrated diffuse white would be 200 nits, 140 nits, 100 nits and 70 nits when shown on display.
Observers

20 observers participated in experiment II, 10 expert observers and 10 naive observers. 10 expert observers were graduate students / faculty members from imaging science or color science.

Observer’s Task

The same five-level methodology was used in experiment II. The same five levels were used. As in experiment I, observers were asked to pick two levels out of five. The five levels were also scaled by the filters. All observers were instructed to the difference between diffuse white and specular white/highlight white before the experiment.

5.2.3 Experiment III

The aim of experiment III is evaluating the impact of the measuring methodology. Therefore, two different methodologies were adopted. In addition to the five-level method in experiment I and II, a single patch with method of adjustment was also used in this experiment. Figure
5.4 shows the configuration of this experiment. Due to the average picture level limitation of the TV, only 25% of the whole TV was used for presenting the image, and the remaining area was set to black. Also the example in Figure 5.4 is an example of the single patch method, where one adjustable patch was added on the center of the image. Observers were sitting in a dark room during the experiment. Observers were seated about 120 cm from the TV, during the experiment.

Images

In this experiment, only images #1, #2, and #4 were used. For the five-level method, the absolute luminance levels of the five patches are 195.9 nits, 255.2 nits, 335.8 nits, 423.3 nits, and 573.9 nits. For the single patch method, 10 levels were provided: 150 nits, 175.8 nits, 205.8 nits, 240.8 nits, 281.9 nits, 330.6 nits, 386.8 nits, 452.5 nits, 529.0 nits, and 598.6 nits.

Figure 5.4: Configuration for experiment III.
Observers

Only expert observers participated in this experiment. Results from experiment I and II showed not much difference between expert observers and naive observers. Therefore, graduate students / faculty members from color science were recruited for this experiment.

Observer’s Task

All observers participated in both methods. For the five-level method, the observers were asked to pick two levels out of the five the same as in experiment I and II. For the single-patch method, observers could adjust the level closest to their estimations of the diffuse white.

5.3 Result and Analysis

All the results and analysis of the three experiments are presented in this section. In experiment I and II, the average result and a comparison between expert observers and naive observers are analyzed. In experiment III, comparison between different methods is demonstrated.

5.3.1 Result of Experiment I

Since it is well known that the perceived lightness is more linear to the log scale of the absolute luminance levels, all the plots and analyses will be on a log scale of the absolute luminance level. This corresponds with the span of the five levels, which are almost linear in log scale. Figure 5.5 plots the overall average result of experiment I. For each image, there are four corresponding data points of different peak luminances, 1000 nits, 500 nits, 300 nits and 200 nits as shown in legend. The five solid lines represent the five provided patch levels. Each data point is plotted with plus-and-minus one standard error bar.

Firstly, it can be seen that the variance across the six images is bigger than the variance across 4 different peak luminances. For images#1, #2, #3, and #6, the data of all four
Figure 5.5: Result of experiment I. The vertical axis is the absolute luminance in log scale. Gray, red, green, blue and magenta solid lines represent the five levels. Dash line represents the calibrated 200 nits diffuse white level.
different peak luminances are within each other’s one standard error. For image#4, result of 1000 nits peak luminance is higher than that of 300 nits and 200 nits peak luminances. For image#4, there is not an obvious white object in the image, which means there is no white reference when observers made estimations. Therefore, a bit larger variance of this image than the others is reasonable. For image#5, the estimated diffuse white of 200 nits peak luminance is lower than that of 500 nits and 300 nits peak luminances. In image#5, the white sweater includes the white objects with a gradient luminance changing. The white object is used by the observers as reference when the observers made their estimations. The texture/detail of that is affected when the image was clipped into different peak luminance levels. When used as a reference, the white object losing texture would have an impact on the estimation. This assumption would need more experiment to verify.

Figure 5.6: Comparison between 10 expert observers’ result (left) and 11 naive observers’ result (right) of experiment I.

The 21 total observers were categorized into expert observers and naive observers. Comparison between the two groups is made to determine if any expertise would affect the result. Figure 5.6 plots the result of 10 expert observers on the left and that of 11 naive observers on the right. Firstly, results of both group showed agreement with the overall group result that the variance across the 6 images is larger than the variance of 4 different peak luminance
levels of the same image content. Therefore, this conclusion, stability of diffuse white across different peak luminances, is valid regardless of expertise. Moreover, the trend across images is similar between the expert observers and the naive observers, as well as that in Figure 5.5. However, the expert observers showed slightly smaller variance over different peak luminance levels in images #5, and #6 than naive observers. The variance across different peak luminance levels in image #4 is larger of the expert observers than that of the naive observers. Therefore, no significant difference can be found between the expert observers and the naive observers. This suggests that expertise of the "diffuse white" would not necessary keep the result more consistent.

5.3.2 Result of Experiment II

Firstly, it should be noted that in this experiment II ND filters scaled whole images. Therefore, the image content did not change but the luminance of the image changed with different filters. Figure 5.7 shows the overall average result of experiment II of the 20 observers. The vertical axis is log scale of the normalized luminance level, which was normalized by the peak luminance. Again, the five solid lines represent the five patch levels. Each data point is the mean of the group chosen diffuse white levels and error bar is one standard error. It can found that variance across images is bigger than the standard error within each image. This agrees with the conclusion from experiment I. Also the general trend of chosen diffuse white across images agrees with that from experiment I, highest in image #3, followed by images #1, #2, #4, then #5, and #6. This again supports that the estimated diffuse white is more relatively an image-based choice. The standard errors of all different ND filters of all images are very close. This standard error is very close to that from experiment I as well.

In this experiment, 20 observers can be divided into two groups: 10 expert observers and 10 naive observers. Figure 5.8 showed the comparison between the two groups. Comparison showed similarity between the expert observers and the naive observers in general trend, and that variance across 6 images is larger than that within the images of 4 different ND
Figure 5.7: Result of experiment II. Vertical axis is the log scale of normalized luminance level. Gray, red, green, blue and magenta solid lines represent the five gray patch levels. Dash line represents the calibrated diffuse white level, 20% of the peak luminance.
filters. For different image contents, the expert observers showed even larger variance across different ND filters on image#5 than the naive observers, and slightly smaller on the rest images. The mean estimated diffuse white levels of the expert observers are also very close to that of the naive observers for all images.

Figure 5.8: Comparison between 10 expert observers’ result (left) and 10 naive observers’ result (right) of experiment II.

### 5.3.3 Result of Experiment III

Figure 5.9 plots the results of 9 expert observers. The blue line and the blue markers indicate the result of the five-level method, the red line and the red markers indicate the result of the single patch method. All markers were also plotted with one standard error. The vertical axis is the absolute luminance in log scale. Both the mean and the standard error were calculated in log scale. Firstly, it can be found that the results of the two methods showed similar trends across images, that the result of image#1 are close to that of image#2 and significantly higher than that of image#4. However, in terms of the absolute luminance level, the result of the five-level method is higher than that of the single patch method. The absolute luminance level is around 350 nits for images#1 and #2, and 300 nits for image#4. For the single patch method, for all three images the absolute luminance level is around
Figure 5.9: Comparison between result of single-patch adjustment methodology and the five-level methodology for the image #1, #2, and #4.
70 nits lower than that from the five-level method. The result showed that the absolute
luminance level of the estimation depends highly on the measurement method.

The diffuse-white level of all the three images is calibrated to 195 nits when presented
on this TV. That was 200 nits for experiment I. They are very close. So it is reasonable to
compare the results of this five-level method with the result from experiment I directly. It can
be found that the mean value in this experiment is slightly lower than that from experiment
I. This could be caused by the difference between the provided absolute luminance levels
in the two experiments. In experiment I, the brightest patch is much brighter than that in
this five-level methodology. The simultaneous contrast could impact the overall perceived
lightness of the five levels if different five levels were provided. Also the standard error of
this experiment in Figure 5.9 is much smaller than that from experiment I (Figure 5.5). This
is more due to the fact that the range of the five levels in experiment III is much smaller
than that from experiment I. Therefore, the standard error is smaller as the chosen levels
were closer to each other.

Additionally, for the single patch method there are three different sizes of the patch,
which are twice of that in the five-level method (size#1), four times of that (size#2) and six
times of that (size#3). The comparison between the three sizes and the five-level method is
plotted in Figure 5.10. Firstly, for all three different sizes, all the mean data are lower than
that of the five-level method. Within the same image, more variance across different sizes is
found for image#2 than image#1 and #4. This is an interesting result. With the patch size
increasing, the patch is covering more and more the white shirt of the lady in image#2 (see
Figure 5.11) while not covering any white reference in image#1 or image#4. According to
the feedback from the observers, their priority in making estimations/decisions are using a
white object as reference. Therefore, covering the white object in the scene would result in
a larger variance in image#2. This is also an attempt of exploring the area-lightness effect
in complex images from Gilchrist [71].
Figure 5.10: Comparison of result of single-patch adjustment with three different patch sizes. size#1 is twice of that in five-level as in Figure 5.3, size#2 is four times of that, and size#3 is six times of that.

Figure 5.11: Preview of original image#2 (top left), with size#1 patch (top right), with size#2 patch (bottom left), and size#3 patch (bottom right).
5.4 Discussion & Conclusions

Observers’ estimations of the diffuse white on the HDR images were explored in three experiments. In experiment I, only the highlights/bright high chroma colors were clipped while the most areas were kept the same. In experiment II, each image was scaled by the ND filters. The ND filter scaled the whole image. In both experiments results, the mean data varied more across the image than different peak luminances or ND filters. Also the general trend across the six images agrees with each other in the two experiments. $R^2$ between the mean of experiment I and experiment II is 0.79. This validates that the observers’ estimation is more image-based than peak luminance or scales. Moreover, the observers were categorized into two groups: expert observers and naive observers. Almost no difference be found by between the two groups in both experiments. However, the expert observers’ result showed slightly smaller standard error than that of the naive observers. Expert observers’ estimations were slightly higher than naive observers.

The result from experiment I showed that clipping the peak luminance, keeping most matte intact, has very limited impact on observers’ estimation of diffuse white level. This tells that observers’ estimation depends more on the matte objects, which corresponds with feedback from observers about their strategies. Results from experiment II showed that observers can adapt completely even 0.3 ND filter, which has 60 nits calibrated diffuse white level. In experiment III, the comparison between different methods demonstrated that the absolute level of observers’ estimation changes with measurement methods. The result of the single patch method is almost 70 nits lower than that from the five-level method, and it is closer to the calibrated diffuse white level 200 nits. The big difference between the single patch method and the five-level method can be explained by the simultaneous contrast effect [72]. Very bright patch in the five-level method makes the perceived lightness of the rest patches much lower compared with the single patch method. So, if the probable measurement methodology, size location and texture of the patch, is used, there is a high possibility that
the estimation can be even closer to the 200 nits. For choosing a diffuse white level in image appearance calculation/color gamut volume, a constant level, the calibrated diffuse white level if the image/video has, can be used, even recommended as a standard for industry practical usage. Due to the peak luminance limitation of the displays, in experiment I and II, the maximum peak luminance is 1000 nits and the peak luminance is around 700 nits. As for the displays having a higher peak luminance, the conclusions would need a verification experiment.
Chapter 6

‘Perfect’ Black On Display

6.1 Introduction

Displays have gained popularity in recent years, especially as high dynamic range (HDR) displays have become widely used and promoted by manufacturers. HDR displays usually demonstrate a higher peak luminance level and a deeper (darker) black level compared with traditional standard dynamic range (SDR) displays. Daly et al. reported an interesting study about human preference of the dynamic range over the shadows, white, and highlights [73]. The experiment showed that a black level darker than 0.001 cd/m² and a peak white of at least 3000 cd/m² are required to satisfy the 90th percentile of observers. Undoubtedly, this study provides guidance for the manufacturer about the necessary dynamic range of displays for customers. It should be noted that this reported dynamic range is for the overall range of different types of images, but not for a single image. Black level gained a great improvement with the development of OLED display technology, which is an active emissive display free of back lighting system flare. OLED is capable of reaching almost 0 black level, a perfect black. However, with the high luminance of the display, more flare would be generated within the optics of the human visual system. Therefore, the glare would be expected to ease the requirement for black level of a display. Murdoch and Heynderickx
published an interesting study about veiling glare over the perceived black of an HDR display [74]. Their results show that the veiling glare from a bright light source would result in the black detail visibility impairment, effectively raising the threshold from 0.0003 cd/m² to 0.0092 cd/m². The threshold was successfully modeled by combining the CIE glare model and the adaptation luminance as a spatially weighted function. The general function is of $1/\theta^2$, where $\theta$ is the visual angle [75]. Later, Stokkermans and Heynderickx extended the model to include the effect of the bright light source over temporal dark adaptation [76]. In addition to the impairment of black detail visibility, the bright light source also diminished viewers’ temporal dark adaptation. In 2016, Stokkermans et al. published another study about effect of different backgrounds on dark adaptation [77]. The results showed a stronger effect from local luminance. Based on the experimental data, Stokkermans proposed a simple model based on that from Murdoch [74], comprising a linear function of the veiling glare. Both of these models are more appropriate for a single glare source. In 2012, Choi et al. [78] proposed an image-based veiling glare model by extending the equation from Stiles and Holladay, $y = k * E/\theta^n$, where $E$ is the luminance from each position, $\theta$ is the visual angle between the source and fixation, and $k$ is a constant scalar. The CIE recommended $n = 2$, while actually that varies, i.e. Stiles found best fitting $n = 1.5$. Choi et al. published another evaluation of the performance of the image-dependent model with the experimental data [79]. Choi et al. derived the $n$ as close to 1 for the background between 20 and 60 gray scale. This could be a limitation about this study that the background itself does not include enough dynamic range. McCann and Vonikakis presented a program calculating the retinal contrast from the scene luminance map based on the glare spread function (GSF), recommended by CIE [80]. The model predicts the lightness scale well within 3.5 log10. However, the simulation of the model also has a limitation of 5.4 log10. Beneficially, this is an image-based model, which will be used for glare calculation in this experiment.

In this study, a psychophysical experiment was designed and conducted to explore the human visual sensitivity to display black level with backgrounds consisting of several different
real images and two designed patterns. This study would enrich the black level experimental data, and the result shows some correlation between the sensitivity and the glare estimation. It could be useful for predicting human black sensitivity for any specific luminance map, or the image data with a display colorimetric model.

![Figure 6.1: Original images used for the experiment.](image)

### 6.2 Experiment Setup & Procedure

#### 6.2.1 Experimental Images

The experiment was set up to measure human’s sensitivity to display black level in comparison to a physical black sample. The physical black is the Metal Velvet black from Acktar\(^1\). The material has almost 0 specular reflectance, well below 1%, with around 1% hemispherical reflectance.

Figure 6.1 shows the original images used in the experiment. It includes four different pictorial images from a HDR image database from Germany [81]. The *firework* stands for a relative dark image. The *car* is a representative of a relative uniform background with a strong highlight spot, the sun in the image. The *lady1* and *lady2* are included to demonstrate

\(^1\)https://www.acktar.com/product/metal-velvet-2/
the effect of the spatial distribution. The two are identical except a flip along the center. All these images were made symmetrical on purpose to prevent any possible asymmetrical flare on the central black circle stimulus. For the firework, car, lady1, lady2, three different exposure levels were used: half the original, the original, and twice the original. For the donut, three different luminance levels were provided: 0.5, 0.75 and 1 in grayscale. For the ring, three different diameters were used: 5°, 8°, and 10° in visual angle around the center of the TV. This set of ring images was focusing on spatial dependency of the image luminance.

Each of these 18 experimental images was manipulated using 6 different black levels for the central circle of each image, making a total of 108 trials. In total, 30 observers participated in this experiment, including four in the pilot test. It should be noted not all the observers finished the whole 108 trials since there was a few breakdowns of the program. On average, around 24 observers completed all 108 trials for which data was collected through the experiment.

6.2.2 Experiment Setup

Observers sat 140 cm from an LG OLED 55-inch C8PUA TV for the experiment, the distance necessary to make the rings appear the desired visual angle. In front of the center of the display, the physical black sample was mounted to a servo motor capable of setting the physical sample into either the left or the right half of the central black circle. A second servo motor moved a black circular card, which was used to cover the central black circle preventing the observer from seeing the physical black sample while it was moving. When the physical black sample was in place for an observation, the circular card moved out of the way, and the observer was asked to indicate which side the physical sample was on – left or right – using a keyboard. For each trial the observer was asked to observe for at least 10 seconds before making a decision, but not longer than 15 seconds. The 10 second was designed to ensure the observers enough observation time, and the 15 second was adopted to prevent a possible strong after-image.
6.2.3 Real-time Display Black Luminance Measurement

Preliminary measurements indicated minor instability of the TV on the desired black level. The variance of the black level could reach 0.02 cd/m\(^2\). Therefore, a real-time black level measuring system was required for the experiment. An Adafruit TSL2591 High Dynamic Range Digital Light Sensor was used to measure the black level at each observation throughout the experiment. Thus, in this experiment, the six black levels used for the center circles in the images were not fixed in absolute luminance.

A Konica Minolta CS2000A was used to derive the relationship between the luminance and TSL sensor readout. The Konica Minolta CS2000A was used to measure the black level directly with a black tube blocking any possible flare. Meanwhile, the TSL sensor data were recorded. For each black level, 10 TSL sensor measurements were taken. Therefore, for each TSL sensor value, there is a mean with a standard error. Figure 6.2 shows the derived function of converting sensor readout to the luminance \( Y \). The plot is in log-log scale. The whole conversion included two linear functions, \( 1 < \log_{10}(TSL) <= 1.5 \) and \( 1.6 <= \log_{10}(TSL) <= 4 \), with a cubic interpolation for \( 1.5 < \log_{10}(TSL) < 1.6 \).

6.2.4 Luminance of the Physical Sample

The physical black sample was only illuminated by the small amount of flare light originating in the TV and reflecting around the room, which was covered in black curtains. Thus, an indirect method was used to measure the luminance of the extremely-dark physical black sample. The idea was to measure the luminance reflected \( (Y_w) \) from a known high-reflectance \( (R_w) \) white object. With the spectral radiance data \( (Spec_w) \), we can infer the flare \( (S_f) \) onto the sample location, and then calculate the luminance reflected from the black physical sample. Eq. 6.1 and 6.2 are used in calculating the \( Y_w \) and \( Y_{\text{sample}} \), respectively. The manufacturer-reported 1\% reflectance of the black sample was used. The ratio \( Y_w/Y_{\text{sample}} \) was found to be very stable in all experimental images, around 128. Therefore, dividing
Figure 6.2: The function of converting TSL sensor readout to the luminance \( Y \) (in log-log scale). The red line is the mean with green dash line indicating the 95% confidential interval. The measured \( Y_w \) by 128 yields \( Y_{\text{sample}} \). Also, the measured \( Y_w \) was very stable for the same images with different central black levels. The results of these physical black luminance levels are presented as black dots in Figure 6.5. Considering the measurement includes some stray light from out of the field of view (FOV), the actual black level should be even lower than the reported value.

\[
S_f \times R_w \times Y_{\text{match}} = Y_w
\]  

(6.1)

\[
S_f \times 1\% \times Y_{\text{match}} = Y_{\text{sample}}
\]  

(6.2)
Figure 6.3: Configuration for measuring the $Y_{\text{white}}$, the calculating the $Y_{\text{sample}}$.

6.3 Experimental Result

6.3.1 Example of Experimental Result

For each of the 18 images, a threshold for the visibility of the black physical sample over the varied black level was found. The threshold was calculated by fitting a psychometric curve (normal CDF) from the guess rate (0.5) to correctly-identified left or right position (1.0). Because the TV black level was not stable, each trial from each observer is considered as a single binomial response, either correct (1) or wrong (0), at its corresponding luminance level. Figure 6.4 shows an example of one experimental image with the log10(Y) on horizontal axis and the proportion of correct responses on the vertical axis. Each red cross is one response with the green star as the initial threshold and the blue line represents the best-fit psychometric curve.

The blue squares in Figure 6.5 are the derived thresholds in absolute luminance, labeled as “Luminance(TSL)”. The physical black sample luminance levels were also plotted as black
stars in the same figure. The horizontal ticks are the image contents. As described in the introduction, each original image generated three variants, half the original exposure as 1, original exposure 2 and twice the original exposure 3. For the ring, three different diameters were used instead of different exposures. The *donut* set is slightly different as well; the three donuts #1, 2, 3 are 0.5, 0.75 and 1 in grayscale. The vertical axis is the luminance in log10 scale.

Mostly, the thresholds are one order of magnitude higher than the luminance from the black sample. This is far higher than the human sensitivity to the sample luminance level. Therefore, the glare was suspected to be the source causing such a threshold. In the next section, the estimated glare will be discussed and analyzed.
6.3.2 Glare Analysis

A program provided by McCann and Vonikakis [80] was used to estimate the retinal glare. The program is mainly a convolution of the scene/image luminance with a kernel, which varies with the age, pupil eye, etc. The convolution kernel was generated based on the CIE recommendation model. Age was set as 25, 30, 40, 50, as well as for two different pupil settings. Not much difference in the estimated glare was found among these settings. Therefore, one estimation of glare from these simulations is used for the analysis. Figure 6.5 shows the estimated glare from the model as the orange dot, along with the experimental threshold, and the luminance from the physical sample. First, the trends along the 18 images are consistent among the three variants, except the glare for ring. The glare model includes a viewing angle-dependent weight function, which decreases dramatically from the center to the periphery. The ring has a quite large diameter, where it results in a very low weight. Therefore the glare prediction is very low. For each subset of images, the firework series
shows nice agreement between the estimated glare and the experimental results. The lady1 experimental results are higher than the glare while the experiment results of the rest are lower than the estimated glare.

Figure 6.6: Glare and APL are plotted against the experimental threshold. Both axes are in log scale. The dash ellipse includes three obvious outliers for the glare against the threshold.

Additionally, the correlation between the estimated glare and the threshold is analyzed. Figure 6.6 demonstrates the glare vs. threshold and average picture level (APL) vs. threshold in log10 scale. Obviously, the glare has much better correlation with the threshold, with three outliers (black-dash ellipse). The three outliers are the ring series. For the ring series, the physical black sample luminance is much higher than the estimated glare. The total luminance level on the retina should be the sum of the physical black sample luminance and the estimated glare. For most images, the glare is the main source on the retina while the physical black sample luminance is the main source for the ring series. The green squares, labeled as "Glare+bk", in Figure 6.6 show that the sum for the ring series are not such
severe outliers in the plot.

<table>
<thead>
<tr>
<th></th>
<th>$R^2$</th>
<th>Threshold</th>
<th>Threshold+bk</th>
</tr>
</thead>
<tbody>
<tr>
<td>APL</td>
<td>0.158</td>
<td>0.183</td>
<td></td>
</tr>
<tr>
<td>Glare</td>
<td>0.245</td>
<td>0.243</td>
<td></td>
</tr>
<tr>
<td>Glare + bk</td>
<td>0.625</td>
<td>0.635</td>
<td></td>
</tr>
<tr>
<td>Glare (no ring series)</td>
<td>0.856</td>
<td>0.880</td>
<td></td>
</tr>
<tr>
<td>Glare + bk (no ring series)</td>
<td>0.857</td>
<td>0.882</td>
<td></td>
</tr>
</tbody>
</table>

Table 6.1: $R^2$ between threshold, threshold+bk and different parameters.

The $R^2$ is calculated for different combinations. There are two different variables in the columns, threshold and threshold+bk. There are five different cases in the rows; APL, Glare, Glare+bk, and Glare, Glare+bk excluding the ring series. Table 6.1 lists all the $R^2$. Clearly, APL has very poor correlation with the threshold. Adding "bk" to "Glare" improves the correlation dramatically. Excluding the ring series achieves a much higher $R^2$, which indicates the ring series is still an outlier among these images. Adding "bk" to the measured threshold did not show too much improvement. Eq. 6.3 shows the best fit between the threshold and the glare (excluding the ring series), and Eq. 6.4 is the best-fitting between the threshold and the (glare + bk) (excluding the ring series). It is reasonable for general images to exclude the ring series, as the black level, rather than glare, dominates.

\[ S = \text{Glare}^{0.464}/11.85 \]  
\[ S = \text{Glare}^{0.481}/11.63 \]
6.4 Summary

Black level luminance sensitivity is highly dependent on glare from image backgrounds. Overall, the experimental data and the correlation with the estimated glare demonstrated the threshold is predictable with a known luminance image. However, if the glare is very low, e.g. the ring series, the correlation failed as the black sample is not a “perfect” black. For that case, observer may be able to adapt to very low luminance level depending on the given time. That would be close to the minimal luminance level the human visual system can detect. For practical usage, Eq. 6.3 gives an adjusted model for predicting the threshold against the pseudo-perfect black. This could be of great benefit to tuning the black level and black level requirement design. For a practical usage, based on the experimental data, a minimal black level of 0.003 cd/m² would be visual “perfect” black.
Chapter 7

Direct Assessment of Human Visual System Simultaneous Dynamic Range

7.1 Introduction

High dynamic range (HDR) displays have gained popularity in recent years for their better capability of color/luminance reproduction, hence better visual experience. HDR display development started with the traditional prototype as replacing the uniform LCD backlight with a spatially modulated projector [82], evolving to the recent local dimming LCD HDR [17] and OLED HDR display technologies, especially with some advanced material development. The LCD-based display has its advantage of achieving a high peak luminance but limited in the achievable black level, mainly due to some internal reflection and limited minimum transmittance. While OLED displays have the advantage of very pure black, approaching luminance of 0 cd/m$^2$, but with some physical limitations on the peak luminance level. Along with the development of the commercial HDR displays, some questions have appeared, e.g. how much dynamic range is enough as for a display [73], how black would be good enough, and what is the impact of the ambient lighting on the HDR display [83]. There is another question remaining: what is the simultaneous dynamic range when the bright and dark
stimuli are presented simultaneously. More specifically, simultaneous dynamic range is the ability of human vision system in discriminating the details in bright and dark region at the same time. The answer would be of great benefit in some fine-tuning of the high dynamic range image, which restricts the adaptation, as well as in devices and algorithms for image capture, processing, and display.

It is known that the human visual system has the capability of adapting to 14 log units of luminance level. From $10^{-6}$ to $10 \text{ cd/m}^2$ is called scotopic range, where the light transduction is mediated by rods, and from $0.01$ to $10^8 \text{ cd/m}^2$ is called photopic range, where the cones are active [84]. The overlapping range is referred to as mesopic vision, where both rods and cones are active. However, it requires a long time, i.e. at least 10-20 minutes, for human to adapt to very low luminance level. Therefore, in most practical usage, the cones are more important to consider, which is still quite large dynamic range compared with most displays.

Two mechanisms contribute to such a large dynamic range of human luminance sensitivity: adapting the cones sensitivity according to the light luminance level, and the intrinsic response of the cone cells has a dynamic range. Variation in pupil diameter also has a role, though it contributes relatively small. It is known that the first mechanism contributes mostly to human visual system’s large dynamic range especially when the average luminance level of the target changes dramatically. The second mechanism, the dynamic range of the photoreceptor cell, can be considered the main determinant of the simultaneous dynamic range although some amount of the rapid local adaptation in the cones does occur. Mainly two reasons limit the simultaneous dynamic range of the cones. Firstly, the optical scattering in the optical media reduces the contrast on the retinal stimulus. Secondly, there is optical response limitation on the photoreceptors and the high order visual mechanisms. There are two different ways of measuring the simultaneous dynamic range: the physiological measurement of the excitation of the photoreceptors and the psychophysical measurement of the simultaneous dynamic range in the perspective of how much dynamic range human can discriminate with a stable adapting level. A range of 2 to 3.5 log units has been reported
as the simultaneous dynamic range via physiological measurement of the excitation of the photoreceptors. Myers reported 2 log units of simultaneous dynamic range [85], Purves and Lotto presented a value of 3 log units [86], and Norman showed a stable value of 3.5 log units in regardless of the background [87]. There is quite a large discrepancy, largely due to the measurement technology and the lack of clear definition of the simultaneous dynamic range. The most recent study of assessment of the simultaneous dynamic range was reported by Kunkel and Reinhard, who used a psychophysical methodology [88]. Kunkel and Reinhard measured the observers’ ability to discriminate a certain contrast level above and below the adapting level separately. Those were considered as the upper and lower bounds of the cone response curve at a certain adapting level. The ratio between the upper and lower bounds was considered as the simultaneous dynamic range. Kunkel reported a maximum of 3.7 log units of simultaneous dynamic range, which is the higher than most previous studies. There are two reasons why they found a higher value: (1) separate measurement of the upper and lower bounds allows the shifting of the adapting level, which would overestimate both bounds in some degree; and (2) the separate measurement of the upper and lower bounds underestimates the effect of the glare on the lower bound, mainly caused by the upper bound. Therefore, 3.7 log units is expected to be an overestimated value compared with direct measurement of the simultaneous dynamic range.

In this study, we measured the simultaneous dynamic range directly with an bright-dark spatially-alternating pattern. The effect of the bright stimulus luminance level and the effect of the stimulus size were explored as well. The comprehensive experimental data were used to build a mathematical model to predict the simultaneous dynamic range. Furthermore, the impact of the spatial frequency of the contrast pattern was explored as well.
7.2 Experiment Setup

7.2.1 Experiment Design

The experiment was designed to measure the simultaneous dynamic range directly. Therefore, the experimental image was designed to contain both bright stimulus and dark stimulus areas. In measuring the simultaneous dynamic range, an edge-blurred Gabor pattern in dark stimulus is used as the discrimination criteria. The edge-blurred Gabor pattern was designed according to mathematical equation as in Eq. 7.1, where $L_0$ is the mean luminance level, $C$ controls the contrast level of the Gabor pattern, $x$ is the horizontal location, $x0$ is the horizontal location of the center of the stimulus, $t$ is the horizontal length of one full cycle, and $S$ is a edge-smoothing function shown in Eq. 7.2, a rectangular function multiplying with a high order ($n=5$) cosine function. The rectangular function is a cut-off function with a diameter of $2 \times r0$, and $(x0, y0)$ is the central position of the Gabor pattern. Eq. 7.1 can also be used to create an edge-smoothing uniform area by setting the $C$ to 0.

$$Y(x, y) = (L_0 + C \times \cos(2\pi \times \frac{x - x0}{t})) \times S$$  \hspace{1cm} (7.1)

$$S = rect(\frac{r}{2 \times r0}) \times 0.5 \times (1 + (\cos(\frac{r}{r0} \times \pi))^n)$$  \hspace{1cm} (7.2)

$$r = ((x - x0)^2 + (y - y0)^2)^{1/2}$$

![Figure 7.1: Illustration of experimental image.](image)
In this experiment, a 2AFC (two alternative forced choice) procedure was adopted. The observers were asked to find the Gabor pattern among the experimental image. Figure 7.1 shows two examples of the experimental image. Both are bright dark spatial-alternating images with a flip of the bright and dark positions. Inside the gray box, there are two identical bright stimuli and two dark stimuli with the same mean luminance level. One of the dark stimuli (top right in the left example, top left in the right example) inside the gray box contains a Gabor pattern, and the other dark stimulus is an edge-smoothed uniform area. Outside of the gray box are just bright stimuli with the same luminance levels, the remaining of which was just black (close to 0 cd/m$^2$). The observer’s task was to find the dark stimulus containing the Gabor pattern. The observer could indicate the choice through left-arrow or right-arrow on the keyboard. It should be noted that the variance of the Gabor pattern was kept on horizontal direction during the whole experiment. Moreover, through the whole study, 10% contrast in dark region was set as the constant test level, i.e. 0.1 log unit in dynamic range. Therefore, $C$ in Eq. 7.1 was set as $5\% \times L_0$, resulting in 10% contrast Gabor pattern. The examples in Figure 7.1 are not actually 10% but have greater contrast to illustrate the experimental image configuration.

### 7.2.2 Apparatus

An Apple Pro XDR display was used for the experiment. The Apple XDR display is an LCD with a spatially-modulated back-light controlled by 576 individual LED zones [89]. The display is capable of reaching a peak luminance of 1600 cd/m$^2$ and 1000 cd/m$^2$ sustained full-screen luminance. The display has a size of 71.8 cm by 41.2 cm with a resolution of 6016 by 3384 pixels. The display was controlled by software developed in Xcode using Objective-C. The software could apply and display 10-bit precision data. The display allows a setting of a diffuse white level, which was set to 50 cd/m$^2$ and the peak luminance was set to 1600 cd/m$^2$. Also, in order to reach the 1600 cd/m$^2$, the experimental image was placed the central 80\% $\times$ 80\% area of the panel. The remaining area of the panel was set as 44.28
Since this is a LCD display, there is image content depending internal reflection. Especially for the dark level, the internal reflection became significant. This is more severe in our experimental image, where the bright stimulus alternates with the dark stimulus. Therefore, there is no standard colorimetric model in characterizing the display while measurements were taken for different settings of the bright stimuli. A CR-100 tristimulus colorimeter was used to measure both the bright stimulus and the dark stimulus. CR-100 was integrated with a large-size sensor for a better precision quick-time in measuring low-luminance stimulus. CR-100 is capable of measuring a range of 0.0007 cd/m² to 5140 cd/m² with a maximum exposure time of 20 seconds [90].

The experiment was conducted in a totally dark room with covered black cloth covering the walls, preventing any possible reflected flare onto the panel. Also two black foam boards were placed on the side of the panel as an addition step to ensure a non-flare environment. The table was covered with black cloth as well. Figure 7.2 illustrates a top view of the experiment setup. The viewing distance was not constant for the experiment, as described in the following section.

### 7.2.3 Stimuli

Due to the limitation of the Apple XDR display, as a back-light LCD display, the smaller the stimulus size is, the larger the internal reflectance and leaked light it has, resulting in a higher minimal luminance level. Therefore, after some testing, a fixed stimulus size was adopted for the entire experiment. While the experimental image has the configuration shown in Figure 7.1, the diameter of the bright and the dark stimulus areas was 118.6 mm, incorporating several backlight zones within each stimulus area.

In this experiment, there were four different luminance levels of the bright stimulus, as 252 cd/m², 452 cd/m², 850 cd/m² and 1600 cd/m². For each bright stimulus luminance setting, the black level was calibrated accordingly to generated Look-Up-Tables (LUTs) between the
Figure 7.2: Schematic top view of the experimental setup with the Apple XDR display on the black-cloth covered table along with two black foam boards.
actual luminance and the input digital count. The LUTs were used to map the desired
luminance level to the input digital count, sent to the display. For example, the black level,
0 input digital count, was around 0.0497 cd/m$^2$, 0.1001 cd/m$^2$, 0.2393 cd/m$^2$, 0.2955 cd/m$^2$
for 252 cd/m$^2$, 452 cd/m$^2$, 850 cd/m$^2$, 1600 cd/m$^2$ bright stimuli respectively. Again, in this
study, 10% contrast in Gabor pattern was set as the discrimination criterion, where $C$ was
set as 5% × $L_0$ in Eq. 7.1. The detail of LUTs can be found in Appendix B.

Also, to explore the effect of the stimulus size on the simultaneous dynamic range, four
viewing distances, visually equivalent to changing the stimulus size, were tested as 2000 mm,
3000 mm, 4000 mm, 5000 mm. The diameters of the stimulus were thus equal to subtended
visual angles of 3.4°, 2.27°, 1.7° and 1.36°.

7.3 Experiments & Results

Two experiments were conducted in this study. The first experiment focused on exploring
how the simultaneous dynamic range changes with different luminance levels of the bright
stimulus and different stimulus sizes. The second experiment focused on the impact of the
Gabor pattern spatial frequency on the simultaneous dynamic range.

7.3.1 Experiment I

Experimental Images

For the four different luminance levels of the bright area, $L_0$ was set as the target luminance
level $L_{bright}$, and $C$ was set as 0 (recall Eq. 7.1). For a better precision, the method of
constant stimulus was used in measuring the simultaneous dynamic range. For each bright
luminance level $L_{bright}$, there were at least 6 different levels of $L_{dark}$. Again, $C$ was set as
5% × $L_{dark}$ for constant 10% contrast. The exact number of dark stimulus luminance levels
was determined in a pilot test. Moreover, the luminance levels of these steps vary with the
$L_{bright}$. For each $L_{bright}$ and $L_{dark}$, both types of bright-dark spatial-alternating images, as
shown in Figure 7.1, were created. In each image, one of the two dark stimuli was randomly selected as the Gabor pattern, and the other dark stimulus was set as constant luminance area. The order of all the experimental images was randomized for each observer.

Additionally, for all four different stimulus sizes, i.e. equivalent to viewing distance in this experiment, were measured as well. The luminance level of the bright region $L_{\text{bright}}$ was the same for different viewing distances. But the $L_{\text{dark}}$ varied with the viewing distance accordingly.

Another important parameter was the spatial frequency of the Gabor pattern. According to the most recent work by Wuerger, Ashfra, et al., the human visual system has a maximum contrast sensitivity around 2 cycle per degree (cpd) [91]. Therefore, the primary spatial frequency was set as 2 cpd with appropriate $t$ in Eq. 7.1. But, it is known that the cycle numbers of the sinusoidal pattern would also have an impact on the measured human sensitivity [5, 91]. Therefore, a minimum of four cycles of the sinusoidal pattern was also set. Therefore, for the smaller stimulus size, the spatial frequency was slightly higher than 2 cpd. The final spatial frequencies of the Gabor pattern were 2 cpd, 2 cpd, 2.35 cpd and 2.94 cpd for 2000 mm, 3000 mm, 4000 mm and 5000 mm respectively. The spatial frequency of the Gabor pattern was within 2 - 5 cpd, the traditional peak sensitivity range [5, 92]. For each $L_{\text{bright}}$, 6 - 8 different $L_{\text{dark}}$ levels images were generated.

**Procedure**

Figure 7.3 shows the flow chart of the experiment. The experiment started with one gray image, 1.025 cd/m². The usage of this non-minimal uniform image can prevent the observer taking advantage of the after image of the darkest image to detect the Gabor pattern. During the pilot test, it was found that if the image was set as the minimal black, the after image of which can be used to detect the Gabor pattern. Therefore, a slight above-minimal image was used to preventing this possible “cheating” method. After 5 seconds, the experimental image was presented with a minimal of 10 second observation time, when the system would
show a text label indicating the system is ready for the choice. The observers were asked not to spend too much time on each trial, preventing after images. There were 5 seconds of the dark uniform image before the next experimental image. The order of the experimental images was randomized for each observer.

**Observers**

One observer, OBS1, participated for all the four luminance levels of the bright stimuli and all four different stimulus sizes. OBS1 repeated the experiments around 28 times of all trials. Another five observers participated for two luminance levels, 452 cd/m$^2$ and 1600 cd/m$^2$, of the bright stimulus with three different stimulus sizes, i.e. viewing distances at 2000 mm, 3000 mm and 4000 mm. The five observers repeated in total 20 times of all experimental images. All observers were color normal and had corrected, or natural 20/20 visual acuity. They ages ranged from 25 to 35 years.

**Results**

**Result of OBS1:** Figure 7.4 plots the measured threshold of the 10% contrast Gabor pattern on the left and the simultaneous dynamic range on the right. The threshold is
defined as the minimum dark stimulus background luminance at which 10% Gabor pattern can be detected on 75% of trials (50% stands for random guess, and 100% for absolute detection). Again, each of the thresholds was derived from a psychometric curve fitting of 6-8 discreet measured black levels. The error bar was generated through bootstrapping. Binomial simulations would create a distributions of the thresholds for each \( L_{bright} \) for each stimulus size. The 2.5% and 97.5% percentiles yield the 95% confidential interval. The simultaneous dynamic range can be simply computed as the ratio between the bright stimulus luminance level and the measured threshold, \( \frac{L_{bright}}{L_{threshold}} \). The horizontal axis in both plots is the stimulus diameter size on a log scale. Each line stands for constant bright stimulus luminance level.

Figure 7.4: Summary result of OBS1 with repeat observations. The 10% contrast threshold is plotted on the left and the simultaneous dynamic range is plotted on the right. Error bars represent the 95% confidential interval through bootstrapping.

Clearly, for OBS1, the threshold showed a linear relationship against the log stimulus size for all four bright stimulus luminance levels. The threshold decreases with the increasing stimulus size for all four bright stimulus settings. It is a reasonable result as the larger stimulus size would cause less glare in the visual system and more capability for local adaptation to the dark stimulus area, hence lower threshold. Moreover, the threshold increases
monotonically with the increasing bright stimulus settings. For the same stimulus size, the brighter the stimulus is, the more glare it will cause in the visual system, hence impairing discriminating ability more. Therefore, the threshold will increase with the increasing bright stimulus luminance level.

The simultaneous dynamic range plot on the right showed the linear relationship against the stimulus size for all four bright region luminance levels. Moreover, in general the simultaneous dynamic range increases monotonically with the stimulus size for all four bright stimulus luminance levels. Also, for the same stimulus size, the simultaneous dynamic range increases with the bright region luminance level mostly except for 3.4°. Another interesting finding is that for the smaller stimulus size, the simultaneous dynamic range increases more significantly with the bright region luminance level than that of the larger stimulus size. For example, for 1.36° stimulus size, the simultaneous dynamic range of 850 cd/m² and 1600 cd/m² are both significantly higher than that of 452 cd/m² and 252 cd/m². The simultaneous dynamic range at 452 cd/m² is significantly higher than that of 252 cd/m² as well. While for 2.27° stimulus size, only the difference between 1600 cd/m² and 252 cd/m² is significant. For 3.4° stimulus size, there is no significant difference among the simultaneous dynamic ranges of all the three luminance levels of the bright stimulus. This is indicating that the simultaneous dynamic range of the human visual system is gradually getting saturated with the increasing stimulus size, regardless of the luminance level of the bright stimulus.

Comparison between Result of OBS1 and Average Result: Figure 7.5 plots the comparison between the comprehensive result of the single observer OBS1 with the average result of five naive observers, which will be referred as average observer in the remainder of this paper. The error bar in the plots represent 95% confidence interval through bootstrapping. The data were plotted on log-log axes as well. For the measured threshold on the left of Figure 7.5, the average observer also showed the threshold decreases monotonically with stimulus size for 452 cd/m², while not for 1600 cd/m². For 452 cd/m², the threshold of average observer decreases significantly at the three measured stimulus sizes, agreeing with
Figure 7.5: Comparison between result of OBS1 and the average result of five observers. The measured threshold of 10% contrast is plotted on the left, and the simultaneous dynamic range is plotted on the right. Both horizontal axes are the stimulus size. Line types represent OBS1 and average observer. Line color represents different luminance levels of the bright stimulus. Error bars represent 95% confidence interval.

For the OBS1 data. For 1600 cd/m², the threshold of OBS1 showed monotonically decreasing relationship against stimulus size but not significantly. The average observer did not demonstrate the decreasing against stimulus size for 1600 cd/m². Also, in general OBS1 showed lower threshold, i.e. higher sensitivity, than the average observer though not significant for all.

For the simultaneous dynamic range plot on the right in Figure 7.5, the simultaneous dynamic range of the average observer increases with increasing stimulus size for 452 cd/m² but not for 1600 cd/m². Moreover, the simultaneous dynamic range with 1600 cd/m² bright stimulus is only significant higher than that of 452 cd/m² for 1.7°. While for 2.27° and 3.4°, the simultaneous dynamic range of the average observer is almost the same for 452 cd/m² and 1600 cd/m². This, more significant difference for smaller stimulus size, agrees with that of OBS1. This showed a general trend that the impact of the bright stimulus luminance level over the simultaneous dynamic range becomes larger with the stimulus size decreasing.
7.3.2 Experiment II

Experimental Images

Experiment II focused on the effect of the spatial frequency of the Gabor pattern on the simultaneous dynamic range. It is known that human visual system has the highest sensitivity at around 2 - 5 cpd for achromatic pattern, as a band pass function [5, 92]. According to the most recent study by Wuerger et al., they reported that the human visual system has highest sensitivity at 0.5 - 2 cpd for luminance level around 0.2 - 2 cd/m² [91]. The peak of the sensitivity is stable at around 2 cpd for luminance level between 20 - 7000 cd/m². Therefore, in this section the simultaneous dynamic range of four different spatial frequencies of the Gabor pattern were measured for one stimulus size as 2.27°, i.e. viewing distance as 3000 mm. The four different spatial frequencies were 1.2 cpd, 2 cpd, 4 cpd, and 8 cpd. The Gabor pattern, with fixed diameter as of 2.27°, making ≈2.7, 4, 8, 16 cycles for the four different frequencies. Two luminance levels for the bright stimulus were used, 452 cd/m² and 1600 cd/m².

Procedure & Observers

Experiment II followed the same as the flowchart in Figure 7.3. The order of the experimental images was also randomized as well. Due to the COVID19 pandemic only one single observer, OBS1, participated in experiment II.

Result

Figure 7.6 plots the result of experiment II, the impact of varying spatial frequency Gabor pattern. The horizontal axis is the spatial frequency of the Gabor pattern, and the vertical axis is the simultaneous dynamic range, $L_{\text{bright}}/L_{\text{threshold}}$. $L_{\text{threshold}}$ is measured threshold of the 10% contrast from the experiment. Both axes are in log scale. The error bars represent for the 95% confidence interval, derived through bootstrapping. Again, this is for stimulus
Figure 7.6: Result of experiment II for OBS1, the simultaneous dynamic range was plotted against the spatial frequency of the Gabor pattern for two bright stimulus luminance level settings.
size as 2.27°, i.e. viewing distance as 3000 mm.

Firstly, the result showed that the shape of simultaneous dynamic range as a function of spatial frequency is a band pass curve, similar to the traditional achromatic Contrast Sensitivity Function (CSF). Moreover, the peak simultaneous dynamic range appears around 2 cpd for 452 cd/m² and between 2 - 4 cpd for 1600 cd/m². This is slightly different from the recent study by Wuerger [91]. Wuerger reported the peak sensitivity 1-2 cpd for luminance level below 2 cd/m². Furthermore, there are two clear differences between the two bright luminance levels: 1) the simultaneous dynamic range is significantly higher of \( L_{\text{bright}}=1600 \) cd/m² than that of \( L_{\text{bright}}=452 \) cd/m², except at 2 cpd; 2) the bandwidth of the 1600 cd/m² curve is larger than that of 452 cd/m². Overall, the experimental data showed similarity with a typical CSF curve, within the range of peak CSF, though demonstrating some difference from the most recent study on the low luminance level. Moreover, the spatial frequency dynamic range curve is clearly the bright stimulus luminance level dependent.

### 7.4 Modeling

The modeling goal was to derive a simultaneous dynamic range model as a function of the bright stimulus luminance level and the stimulus size in an allowable range. However, due to the limited data of the average observer, the model was derived based on the comprehensive data of the single observer, OBS1, from experiment I to generate a descriptive model form. The average observer data were then compared with the fitted model. Two different models were proposed to fit the experiment I data. One is linear model in log scale, which is clearly a fitting based on the experiment data. However, there is a fundamental drawback of the log linear model that there is no saturation level. Therefore, in the second model, a saturation level is included, a nonlinear model on the log scale. Both models have advantages and disadvantages. Since, there is no previous modeling, or similar experimental data on the direct simultaneous dynamic range measurement, no comparison with other studies can
be made. Moreover, for experiment II, a log parabola function was used to model the simultaneous dynamic range as the function of the spatial frequency for each bright stimulus luminance level.

### 7.4.1 Experiment I Modeling

Two different models were proposed to fit the simultaneous dynamic range as a function of stimulus size and the bright region luminance level. Both models were based on threshold prediction, which can be easily transformed into the simultaneous dynamic range. The first model is called log linear model, a linear function in log scale as in Eq. 7.3, where \( D \) stands for the diameter of the stimulus in the experimental image, \( DR \) means the simultaneous dynamic range. The constant parameters, \( k \) and \( a \), of the linear function vary with the luminance level of the bright stimulus. The second model is a nonlinear function as in Eq. 7.4, where \( D \) is the diameter of the stimulus as well and \( k, n, a, b \) are constant parameters. This is a modification of the cone luminosity response [93]. These parameters will be derived for each bright stimulus luminance level. Optimization code can be found in Appendix C.

\[
\begin{align*}
\log_{10} L_{\text{threshold}} &= k \times \log_{10} D + a \\
\log_{10} DR &= \log_{10} L_{\text{bright}} - \log_{10} L_{\text{threshold}} \\
L_{\text{threshold}} &= k \times (1 - \frac{D^n}{D^n + b}) + a \\
\log_{10} DR &= \log_{10} L_{\text{bright}} - \log_{10} L_{\text{threshold}}
\end{align*}
\]

(7.3)

(7.4)

Figure 7.7 plots the original experiment data and best-fitting curves using the two methods. Each curve is best fitting for constant luminance level of the bright stimulus. The threshold fitting was plotted on the left and the simultaneous dynamic range was plotted on the right. The parameters of the two methods were listed in Table 7.1. It can be found that, both methods fit the data well. In general, the nonlinear fitting method (dash line) works better than the log linear fitting method (solid line). The log linear fitting works best
Table 7.1: Optimized parameters for log linear and nonlinear fitting over experiment I data.

<table>
<thead>
<tr>
<th>$L_{\text{bright}}$</th>
<th>Method 1 (log linear)</th>
<th>Method 2 (nonlinear)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$k$</td>
<td>$a$</td>
</tr>
<tr>
<td>252 cd/m$^2$</td>
<td>-1.419</td>
<td>-0.286</td>
</tr>
<tr>
<td>452 cd/m$^2$</td>
<td>-1.029</td>
<td>-0.204</td>
</tr>
<tr>
<td>850 cd/m$^2$</td>
<td>-0.676</td>
<td>-0.098</td>
</tr>
<tr>
<td>1600 cd/m$^2$</td>
<td>-0.761</td>
<td>0.135</td>
</tr>
</tbody>
</table>

for 452 cd/m$^2$ and 1600 cd/m$^2$. The nonlinear method fits the mean threshold very well except for 252 cd/m$^2$. For the simultaneous dynamic range plot on the right, the nonlinear method showed advantage of including a saturation level with increasing stimulus size. The nonlinear fitting curves (dash line) showed clearly saturation around 3.4° for 252 cd/m$^2$, 452 cd/m$^2$ and 850 cd/m$^2$ but not for 1600 cd/m$^2$.

**Constrained fitting:** During the examination of the two methods, it was found that the three fitting lines (252 cd/m$^2$, 452 cd/m$^2$, and 850 cd/m$^2$) intersect with each other at almost the same position for log linear method (recall solid lines on left in Figure 7.7).
Therefore, to simplify the log linear model, one constraint was added: the four fitting lines of the four $L_{bright}$ will intersect at almost the same position. Moreover, for the nonlinear method 2 it was found that two of the four parameters in Eq. 7.4 can be set as constant across different $L_{bright}$ while maintaining a good performance. The best performance was found for constant $b=0.2$, $a=0.1$. The optimized parameters of the two methods under the constrains are listed in Table 7.2. The parameters, $k$ and $a$, in method 1 can be simplified by a linear prediction. The number of the parameters in method 2 (Eq. 7.4) can be reduced from 4 to 2. Constrained fitting optimization can be found in Appendix C.

Figure 7.8: Result of experiment I for single OBS1 with constrained fitting. Method 1 fitting (log-linear) is in solid lines, and method 2 (nonlinear) is in dash line.

Figure 7.8 plots the experimental data with the optimized constrained fittings. For method 1 (log-linear) fitting, the fitting lines change most for 252 cd/m$^2$ and 850 cd/m$^2$ with the constrains, but less change for 452 cd/m$^2$ and 1600 cd/m$^2$. While for the nonlinear method 2 fitting, the constrains do not change too much within the experimental stimulus size. For the nonlinear fitting method 2, the constant $a$ stands for the saturation level, which is the threshold for very large stimulus size where the observers could adapt more to the dark stimulus in regardless of the bright region luminance level. Therefore, the threshold would saturate for very large stimulus size. This is the advantage of the nonlinear fitting.
### Table 7.2: Optimized parameters under constrain for log linear and nonlinear fitting over experiment I data.

<table>
<thead>
<tr>
<th>$L_{\text{bright}}$</th>
<th>Method 1 (log linear)</th>
<th>Method 2 (nonlinear)</th>
</tr>
</thead>
<tbody>
<tr>
<td>252 cd/m²</td>
<td>-1.0970   -0.3685</td>
<td>2.9032  2.7433</td>
</tr>
<tr>
<td>452 cd/m²</td>
<td>-0.9761   -0.2140</td>
<td>3.5198  1.7505</td>
</tr>
<tr>
<td>850 cd/m²</td>
<td>-0.8454   -0.0469</td>
<td>4.3984  1.0231</td>
</tr>
<tr>
<td>1600 cd/m²</td>
<td>-0.7144   0.1204</td>
<td>7.8013  1.0035</td>
</tr>
</tbody>
</table>

To compare the performance of the non-constrained optimization and the constrained optimization, the root-mean-square-error (RMSE) was computed for different $L_{\text{bright}}$. The RMSE in log unit, the same for the thresholds and the simultaneous dynamic ranges, are listed in Table 7.3. Overall, the RMSE is below 0.03 in log unit (7% in linear scale), which is very small. With the constrains, the RMSE increases slightly. For 252 cd/m² and 850 cd/m² of method 1, the RMSE increases more under the constrains. The RMSE increases more only for 252 cd/m² of method 2. Overall, the RMSE increases slightly with the constrains but still small compared with the simultaneous dynamic range value, 3-3.3 log unit.

### Glare-based model

As mentioned in the introduction, the glare caused by the bright stimulus is one of the main resource limiting the human visual system in discriminating the Gabor pattern. Therefore, the estimated glare over the dark stimulus is expected to be related with the 10% contrast pattern threshold. For glare estimation, there is a widely used Stiles-Holladay model as in Eq. 7.5, where $\beta$ is the estimated glare, $k$, and $n$ are constant, $E$ is the illuminance from the source, and $\theta_G$ is the disparate visual angle between the glare source and the fixation. Stiles proposed $n = 1.5$ [94], and Holladay reported $n = 2$[95]. McCann showed a
Table 7.3: RMSE in log unit, the same for $L_{\text{threshold}}$ and the simultaneous dynamic range, of the two methods, for no constrain and constrained conditions.

<table>
<thead>
<tr>
<th>Method 1 (log linear)</th>
<th>Method 2 (nonlinear)</th>
</tr>
</thead>
<tbody>
<tr>
<td>no constrain</td>
<td>constrained</td>
</tr>
<tr>
<td>252 $\text{cd/m}^2$</td>
<td>0.0077 0.0305</td>
</tr>
<tr>
<td>452 $\text{cd/m}^2$</td>
<td>0.0111 0.0132</td>
</tr>
<tr>
<td>850 $\text{cd/m}^2$</td>
<td>0.0116 0.0267</td>
</tr>
<tr>
<td>1600 $\text{cd/m}^2$</td>
<td>0.0097 0.0116</td>
</tr>
</tbody>
</table>

Program to compute the retinal contrast image considering the human visual system glare [80]. McCann’s program is based on the CIE standard [96], which adopts the $n = 2$.

$$\beta = k \times \frac{E}{(\theta_G)^n}$$

(7.5)

Figure 7.9 plots the 10% contrast threshold against the mean of the estimated glare over the Gabor pattern area. This is a log-log plot. The error bars are the 95% confidence interval of OBS1 data. The solid lines are best fittings for the four bright stimulus luminance levels following Eq. 7.6, where $k$, and $b$ are bright stimulus luminance dependent constants, $\beta$ is the estimated glare. Surprisingly, the three lines of 252 $\text{cd/m}^2$, 452 $\text{cd/m}^2$ and 1600 $\text{cd/m}^2$ intersect at the same point. Therefore, constraining all the four fittings lines intersecting at the same point would reduce the parameters from 2 to 1 for each line as in Eq. 7.7, where $k$ is luminance dependent, $\beta_0$ (3.43) and $L_0$ (0.051) are constants. The dashed lines in Figure 7.9 are the optimized results under the constraint. All the constants and the RMSE are listed in Table 7.4. It can be found that, there is only slight change of the line of 850 $\text{cd/m}^2$, while the rest almost remain the same. Even under the constrain, all the fitting lines are
Table 7.4: Parameters for the glare-based model: bright stimulus luminance level dependent $k$ and $b$ Eq. 7.6, and $k$ for Eq. 7.7.

<table>
<thead>
<tr>
<th>Luminance level</th>
<th>$k$</th>
<th>$b$</th>
<th>RMSE ($\log_{10}$)</th>
<th>$k$</th>
<th>RMSE ($\log_{10}$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>252 cd/m²</td>
<td>2.168</td>
<td>-2.443</td>
<td>0.0097</td>
<td>2.189</td>
<td>0.0098</td>
</tr>
<tr>
<td>452 cd/m²</td>
<td>1.579</td>
<td>-2.172</td>
<td>0.0102</td>
<td>1.509</td>
<td>0.0123</td>
</tr>
<tr>
<td>850 cd/m²</td>
<td>1.036</td>
<td>-1.671</td>
<td>0.0107</td>
<td>1.254</td>
<td>0.0240</td>
</tr>
<tr>
<td>1600 cd/m²</td>
<td>1.164</td>
<td>-1.952</td>
<td>0.0108</td>
<td>1.126</td>
<td>0.0114</td>
</tr>
</tbody>
</table>

The RMSE is very small in log unit, except for that of 850 cd/m² under the constrain (reaching 0.02 in log unit, 5% in linear unit).

\[
\log_{10} L_{\text{threshold}} = k \times \log_{10} \beta + b \tag{7.6}
\]

\[
\log_{10} L_{\text{threshold}} = k \times (\log_{10} \beta - \log_{10} \beta_0) + \log_{10} L_0 \tag{7.7}
\]

Clearly, the threshold has a positive linear relationship with the mean estimated glare for each bright stimulus luminance level. Moreover, the slope of the fitting lines decreases with increasing bright stimulus luminance level, which means the impact of the glare decreases with higher bright stimulus luminance level. Also, for the same estimated glare level, the threshold is higher for the low luminance level of the bright stimulus (smaller stimulus size) than that of the high luminance level (larger stimulus size).

### 7.4.2 Experiment II Modeling

Log parabola function has been used to model contrast sensitivity function (CSF) for a long time [92, 97–99]. Especially, the most recent work from Wuerger, et al. showed that model
Figure 7.9: Each data is plotted the mean with the 95% confidence interval. The solid lines are the best fittings for the four luminance levels. The dash lines are the optimized lines with constrain.
works well for a high dynamic range from 0.02 cd/m$^2$ up to 7000 cd/m$^2$ [91]. Therefore, a log parabola was used to fit our experiment II data as well, as in Eq. 7.8. $DR$ stands for the simultaneous dynamic range, $DR_{max}$ is the maximum simultaneous dynamic range, $f$ is the spatial frequency, $f_{max}$ is the spatial frequency with the maximum simultaneous dynamic range, and $b$ is the bandwidth.

$$
\log_{10}DR = \log_{10}DR_{max} - \left(\frac{\log_{10}f - \log_{10}f_{max}}{b}\right)^2
$$

(7.8)

Figure 7.10: Best log parabola fitting for 452 cd/m$^2$ and 1600 cd/m$^2$ separately for varying spatial frequency patterns.

Figure 7.10 showed the log parabola fitting separately for 452 cd/m$^2$ and 1600 cd/m$^2$ in dash lines. (The optimization code can be found in Appendix C.) Table 7.5 lists the
<table>
<thead>
<tr>
<th>$L_{bright}$</th>
<th>$\log_{10} DR_{max}$</th>
<th>$f_{max}$</th>
<th>$b$</th>
<th>RMSE ($\log_{10}$ unit)</th>
</tr>
</thead>
<tbody>
<tr>
<td>452 cd/m²</td>
<td>3.24</td>
<td>2.6</td>
<td>4.6</td>
<td>0.0170</td>
</tr>
<tr>
<td>1600 cd/m²</td>
<td>3.40</td>
<td>2.7</td>
<td>6.2</td>
<td>0.0135</td>
</tr>
</tbody>
</table>

Table 7.5: Optimized parameters for log parabola fitting over experiment II data.

Optimized parameters for 452 cd/m² and 1600 cd/m². $DR_{max}$ is higher for 1600 cd/m² than that for 452 cd/m². Moreover, the bandwidth $b$ of 1600 cd/m² is larger than that for 452 cd/m². Interestingly, $f_{max}$ is very close for 452 cd/m² and 1600 cd/m². Overall, the fitting simultaneous of 1600 cd/m² is higher than that of 452 cd/m² in the range between 1.2 cpd to 8 cpd. Obviously, in this experiment data, the $f_{max}$ is larger than the $f_{max}$ from Wuerger’s work [91], 0.5 - 2 cpd. Moreover, the RMSE in $\log_{10}$ unit listed in Table 7.5 showed only 0.017 and 0.0135 of the log parabola fitting.

## 7.5 Discussion

### 7.5.1 Experiment I

Simultaneous dynamic range is an important concept describing human visual system’s capability in discriminating the details when high dynamic range stimuli were presented at the same time. This value has been measured a few times in previous studies, range from 2 - 4 log units. However, none of them measured the simultaneous dynamic range in a direct way through psychophysical methodology. The most recent studies from Kunkel, et al. reported a value of 3.7 log units simultaneous dynamic range through measuring the threshold of a Gabor pattern with a certain contrast level above and below the adapting level separately [88]. This indirect measurement method would overestimate the actual simultaneous dynamic range due to the lack of ensuring stable adaptation and lack of glare consideration. This could explain why the reported value from Kunkel is higher than most previous studies, as well as this study. So far, there is no other direct measurement of the simultaneous
<table>
<thead>
<tr>
<th>Method 1 (log linear)</th>
<th>Method 2 (nonlinear)</th>
</tr>
</thead>
<tbody>
<tr>
<td>452 cd/m$^2$</td>
<td>0.034</td>
</tr>
<tr>
<td></td>
<td>0.005</td>
</tr>
<tr>
<td>1600 cd/m$^2$</td>
<td>0.047</td>
</tr>
<tr>
<td></td>
<td>0.047</td>
</tr>
</tbody>
</table>

Table 7.6: RMSE in log unit, the same for $L_{threshold}$ and the simultaneous dynamic range, of the two methods with constrain for average observer’s data.

dynamic range in the perspective of the observers. No further comparisons with any other studies can be made.

Figure 7.11: Best log-linear and nonlinear fittings for 452 cd/m$^2$ and 1600 cd/m$^2$ separately of the naive observers’ data.

In this study, a comprehensive direct measurement of the simultaneous dynamic range was used to propose a model in predicting the simultaneous dynamic range of 10% contrast in dark stimulus as a function of the stimulus size. Two different methods were proposed as log-linear and nonlinear model. The parameters of both are bright-stimulus luminance level dependent. The log-linear model has its advantage of simplicity. However, it does not include the saturation feature, which theoretically should be included in a comprehensive model for extreme large stimulus size. However, both models are based on one single observer’s data.
Therefore, it is necessary to examine the model with the limited collected average observer’s data. Figure 7.11 plots the examination of the average observer’s data of the two bright-stimulus luminance levels, 452 cd/m$^2$ and 1600 cd/m$^2$ with two constrained fitting models. Table 7.6 lists the RMSE of the two fitting methods over the two luminance levels. In general, the RMSE is very small in $\log_{10}$ unit, no more than 0.047 compared with 3.3 log unit simultaneous dynamic range. Moreover, the two methods showed the same performance for 1600 cd/m$^2$ luminance level in terms of RMSE, higher than that of 452 cd/m$^2$. For 452 cd/m$^2$, method 2 showed almost a perfect fitting, while the RMSE is higher for method 1 fitting. Though the RMSE of the average observer’s fitting is higher than the OBS1’s data in Table 7.3, the examination of the model using average observer’s data promises the generalization of the simultaneous dynamic range model for the 10% contrast, as a function of the stimulus size. The generalization of the model would require more average observers’ experimental data.

**Glare-based model** presented in the modeling works well with OBS1 data. It showed that for the same estimated glare the threshold for higher luminance level (larger stimulus size) is lower than that of the lower luminance level (smaller stimulus size). There are two possible explanations: 1) the pupil size will change with the bright stimulus luminance level and the stimulus size. While the glare model does not take that into consideration. 2) The glare itself is not enough to explain the 10% contrast threshold in this pattern. Moreover, there is a difference between the glare and threshold. Choi and Alabni, et al., proposed an image-dependent model, where the threshold was predicted with accumulated glare from the rest of the image [78]. Choi reported a background-dependent $n$ for a certain region of interest (ROI). Though with only two discrete backgrounds, the result showed an interesting point that $n$ should vary with background, maybe even with specific pattern, because the background and the image content would affect the state of the adaptation, maybe the pupil size as well. Though, lacking such research work prevents a further analysis of the glare-based model, Choi’s work here showed the validation of using glare to predict the 10% contrast
<table>
<thead>
<tr>
<th>Luminance level (cd/m²)</th>
<th>RMSE (log₁₀ unit) no constrain</th>
<th>RMSE (log₁₀ unit) constrained</th>
</tr>
</thead>
<tbody>
<tr>
<td>452</td>
<td>0.0128</td>
<td>0.0163</td>
</tr>
<tr>
<td>1600</td>
<td>0.0471</td>
<td>0.0509</td>
</tr>
</tbody>
</table>

Table 7.7: RMSE in log unit, the same for $L_{\text{threshold}}$ and the simultaneous dynamic range, of the fitting results for average observer’s data against the estimated glare (in Figure 7.12). pattern (the simultaneous dynamic range) for the given bright region luminance level and the threshold prediction being background/image content dependent. The bright stimulus dependent glare-based model is consistent with Choi’s work.

This glare-based model was tested with the average observers’ data as well as the other two models. The average observers’ data are fitted with Eq. 7.6, and Eq. 7.7. Figure 7.12 plots the results and Table 7.7 lists the RMSE as the fitting error. The line fits the data of 452 cd/m² luminance level very well, much better than that of 1600 cd/m². Clearly, the constrain does not change the fitting line too much. The RMSE is slightly higher for the constrained fitting, reaching 0.0509 log unit (12% in linear scale) for 1600 cd/m². Clearly, the error is higher for 1600 cd/m², but still it is within the 95% confidence interval. Again, the lack of more average observers’ data prevents the generalized model but the fitting of average observer’s data showed the promising performance of the model.

### 7.5.2 Experiment II

Impact of Gabor pattern spatial frequency was explored and modeled in experiment II. The log parabola function fits the experimental result well, though the log parabola was used for traditional CSF model mostly. Experiment II is slightly different from the traditional CSF measurement, where the sensitivity was measured for varying spatial frequencies at a certain mean luminance level. The threshold of a constant relative contrast level, 10%,
Figure 7.12: Best log-linear fittings without constrain (solid lines) and with constrain (dash lines) for 452 cd/m² and 1600 cd/m² separately of the naive observers’ data.
was measured for different spatial frequencies of the bright dark spatial-alternating pattern. Another critical difference is the adapting time. For luminance level between 0.4 - 2 cd/m² Gabor pattern, it usually requires much longer adaptation time, for example observers spent 5-10 minutes adapting to the luminance level for 0.02 cd/m² and 0.2 cd/m² in Wuerger’s experiment I [91]. However, the bright dark spatially-alternating pattern would prevent the complete adaptation to dark stimulus. Additionally, a minimum of 10 second was set to prevent any rush decision. Mostly observer did not spend much longer than 10 seconds. Moreover, the Gabor pattern has large enough size to ensure enough cycles of the high spatial frequency, e.g. 16 cycles for 8 cpd. This meets the minimal of 7 cycles from Barak’s study [100]. Though Manituk also reported the target size impact over CSF at low luminance level [101], the target size is smaller than that in this study. Manituk’s work can not be compared with our data directly. However, the data from Manituk’s work showed the impact of increasing size (up to 1.5°) decreases. Therefore, for the target size 2.27° in this experiment, we can speculate the very little impact due to the limited target size. Experiment II is representative for the spatial frequency effect on the simultaneous dynamic range.

In this study, the peak simultaneous dynamic range was found around 2.6/2.7 cpd, which is within the range of the traditional peak CSF 2-5 cpd [5, 102]. However, this is slightly higher than the recent study of the low luminance level from Wuerger, where the peak was found no more than 2 cpd [91]. Two facts may contribute to the difference: the low cycles (2 full cycles) in Wuerger’s data, and the observers in Wuerger’s experiment adapted more to that low luminance level as the traditional CSF measurement. The peak simultaneous dynamic range around 2.6-2.7 is a reasonable result, not conflicting with previous reported data.
7.6 Conclusion

In this study, the simultaneous dynamic range was measured directly through the bright dark spatially-alternating pattern for different stimulus sizes with varying bright stimulus luminance level. The simultaneous dynamic range was found to be bright area luminance-level dependent. Within 1600 cd/m² bright-region luminance level, the simultaneous dynamic range was found to increase monotonically with the bright-area luminance level but gradually saturated. Moreover, the simultaneous dynamic range was found to increase with the stimulus size. A maximum value of 3.3 log units for the average observer, 3.47 for OBS1, was found for 1600 cd/m² bright-area luminance level of 3.4° stimulus size. Two different methods were proposed to fit the experimental data. Both methods work well for the experiment I data. The method 1, log linear, has an advantage of simplicity but not including a saturation level for extreme large size. Method 2, a nonlinear method, fits the data better in terms of RMSE. Moreover, method 2 has advantage of including the saturation level for increasing size but requiring nonlinear parameters interpolation. Fitting average observer’s data with the two methods showed the promising future of generalizing the model for average observers’ data. Very similar conclusion can be found for the glare-based model, which fits the 10% contrast threshold against the estimated glare for a given bright stimulus luminance level. It can be found that the linear fitting works well but the slope and offset of the line are bright stimulus luminance level dependent. It works well for the average observer’s data. All these demonstrated that the simultaneous dynamic range can be predicted with simple model with bright stimulus luminance level dependent parameters. However, lack of average observer’s data prevents a further generalization of the model. Therefore, the major future work would be collecting more average observers’ data and generalize the simultaneous model. Also, in addition to the 10% contrast level, more contrast levels can be explored for a comprehensive simultaneous dynamic range model. Experiment II showed the impact of spatial frequency of the Gabor pattern on the simultaneous dynamic range. The log parabola function, a
band pass curve, models the impact well, similar to the achromatic CSF model. The peak simultaneous dynamic ranges appeared around 2.6 cpd for both luminance levels. This is within the range of the traditional peak CSF range (2-5 cpd). The model is based on one single observer data. The future step of the spatial frequency effect would be generalizing the model with a comprehensive average observers’ data.
Chapter 8

Summary

Along with the development and commercialization of the HDR WCG display, it is important to better understand the HDR display in the perspective of perception and the normal observers. Traditionally, most color perception was focusing on the limited dynamic range of display, or reflective objects. Though there are some developments of the HDR color spaces, appearance models, those are not sufficient. Therefore, it is meaningful to explore the perception of HDR WCG displays.

In this dissertation, four sections about HDR display were explored: 3D color gamut volume, diffuse white level when presented in a HDR display, ‘perfect’ black level on a display, and the simultaneous dynamic range. Color gamut volume is a metric used to characterize the display’s color reproduction ability. Traditionally, the chromaticity area coverage by the primaries of the display was used as the main color gamut metric. However, 2-dimensional chromaticity diagram is not sufficient for describing colors. Therefore, 3-dimensional volume is the least requirement for a better color gamut volume metric. Moreover, a color appearance space would be more meaningful due to its better correlation with color attributes. Therefore, the 3D volume expressed in a color appearance space is a better choice for color gamut volume. The study on the 3D color gamut volume presented that the 3D color gamut volume can be simplified by a multiplication of three factors, associated with the primaries’
chromaticity diagram coverage on $x'y'$, peak luminance, and the RGB/RGBW configuration of the display. Three simple polynomial functions can be used to compute the three factors separately. An independent dataset showed the high accuracy of the model. Additionally, two experiments were conducted to explore the perceptual color gamut volume in the perspective of the observers with different peak luminance capabilities. The experiments followed two different methods of image processing. The two processing methods corresponded with the two diffuse white settings in the model. A fundamental difference was found between the experimental result and the model prediction. The experimental result showed a better correlation with the log scale of the peak luminance while the model predicts a better correlation with the linear scale of the peak luminance. The fundamental difference was found between the model prediction and the experiments. The model assumed the uniform usage of the 3D color gamut volume, all colors are assumed to be used equally. However, in reality most of the HDR image data are below some level, while only some highlights, bright saturated colors are used for specific applications, i.e. highlights of reflective surface, direct lights within the image. Therefore, the model predicts the mathematical 3D color gamut volume well while the perceptual color gamut volume should be considered separately from the mathematical model. To better describe the display’s color reproduction capability, different metrics should be used for different applications/perspectives.

Diffuse white was found to be an important question in both the model and the perceptual color gamut volume. Diffuse white in HDR image was explored in the second part of the dissertation. The diffuse white was found to have very poor correlation with the image average picture level (APL). The estimated diffuse white was found to be stable for images with clipped peak luminance levels. Moreover, for images scaled to different peak luminance levels, the ratio between the diffuse white and the peak luminance level was found to be stable for different scaling factors and different image contents. Moreover, it was found the exact luminance level of the estimated diffuse white varies with the measuring methodologies. Additionally, a white area in the image was helpful in reducing the estimation error.
Furthermore, the size of the matching patch was found to affect the exact luminance level as well. The experiment validates the stable assumption about the diffuse white to a great extent but with some variances. This is very interesting work on computing the image appearance for same content but with different peak luminance levels. This is a pioneering study on the perceptual estimation of diffuse white presented in HDR display for naive observers, different from experts. This work has great value for further development of HDR image appearance with different image processing methods.

The HDR display, especially the OLED HDR display, has its advantage of a deeper black. However, visually there should be a threshold, which is equivalent to ‘perfect’ black, 0 cd/m². A visual experiment was conducted to measure the threshold above the ‘perfect’ physical sample. A comprehensive set of different types of images was used to generalize the conclusion. The experimental result showed an interesting correlation between the threshold and the predicted glare over the target area. Therefore, the ‘perfect’ black prediction model was proposed based on the experimental data. Theoretically, the model can predict the requirement for visually perfect black level. This work can be beneficial to HDR display design, and evaluation of HDR display, especially the black level.

The last section of the dissertation focused on the simultaneous dynamic range measurement and modeling. The simultaneous dynamic range was measured for different peak luminance level for different stimulus sizes. It was found that a close-to log linear relationship can model the experimental result very well. Moreover, the model can be simplified through adding constraints to the model. The model is examined by the independent average observer’s data. The RMSE (no more than 0.05 log unit) is very small compared to the 3-3.3 log unit simultaneous dynamic range. The nonlinear model, a modification of the cone response, has an advantage of introducing the saturation level for large-size stimulus size while it requires more parameters than the log-linear model. Additionally, the glare based model works well with the log-linear fitting. The log-linear fitting requires peak luminance dependent parameters as well. This agrees with some previous studies. This initial modeling
works well with the single observer’s data. The verification by the average observer’s data demonstrates a promising generalization of the model.

The four aspects workings on the HDR display perception aim at exploring and characterizing the perceptual difference between HDR display and SDR display. Color gamut volume is a metric to quantize the displays’ color reproduction capability. In this study, the computational volume was calculated in color appearance models, further a model was proposed. Of course, this is not the only way to propose and compute a color gamut volume. Kunkel and et al proposed the volume disk in ICtCp color space [56]. This is an interesting way to assess color gamut volume claiming the full adaptation to different absolute luminance level. It assumed a complete luminance adaptation, while ICtCp space itself does not include such an adaptation. Kunkel’s adapting hull method is a way of computing the color reproduction’s capability in discriminating the very fine details, close to Just-Noticeable-Difference (JND) level. This is more of micro-volume, which works over around the threshold. However, it will be of dramatically difference from the marco-difference, or called perception scales.

To assess the display system color reproduction capability, the proposed model should be close to the display system’s main application, presenting images and video contents. In presenting the natural images, observers/users would be able to achieve a more complete adaptation to different luminance levels with sufficient time. But they would never achieve to the fine details level, as in Kunkel’s adapting hull volume model. Moreover, for videos, unless the video content is very dark through a long time, most observers could not reach much adaptation to different luminance level in a short time. Therefore, in this case, the adapting hull method is not suitable to generalize the display’s color reproduction ability.

In practice, HDR displays would only show its superiority over some specific types of images or videos. Therefore, the proposed model in computing color gamut volume in color appearance models is closer to the display’s capability in serving its application. Undoubtedly, black level is indeed important property while not reflected much in the proposed color gamut volume. Black level of the display would have a great impact in presenting deep black
when necessary, saying showing some details in dark region in a dim image. But those images would be not the main types of images used by displays. Additionally, for those dim images, unless the observers focus on the very dark region for a long time otherwise, the black level of the display may not have strong impact on visual experience. The work about visually “perfect” black could be of research value and be usage for display manufacturers but it is not the only important point in assessing display’s color reproduction capability. Moreover, peak luminance seems play another important role. The higher the peak luminance is, the brighter/more colorful the display is capable of presenting. Additionally, with a higher peak luminance level, the whole image would generate stronger glare in the visual system, hence impairing human’s capability in discriminating the details in the dark. Therefore, the brighter peak luminance would soften the black level requirement. As presented in the simultaneous dynamic range work, it can be found that the simultaneous dynamic range is very limited to a maximum value of 2000:1. Therefore, the requirement of black level for a daily used display is not as strictly as the minimum requirement in “perfect” black, though it will be beneficial if the display would achieve that requirement. Therefore, the impact of a slight higher black level may be overestimated in mostly images. The work about diffuse white estimation showed that with peak luminance levels from 200 nits to 1000 nits do not have strong impact on diffuse white estimation. Therefore, the difference between 0.1 nits black and 0.01 nits black ,with a 200 nits diffuse white level, is very limited. Again, certainly there will be difference when showing very dim image with fine details. However, for mostly images, the black level may not be that strong impact over display color reproduction capability. Therefore, a general color gamut volume metric in color appearance models to characterizing display color reproduction capability is a reasonable choice for now. Indeed, it requires much more work to validate and optimize the usage of the metric, even with amendments i.e. for different applications using different parameters.
8.1 Future Work

- For the 3D color gamut volume, additional work can be added on the 3D color gamut volume over a higher dynamic range, and a comparison between the static 3D color gamut volume and the adaptive color gamut volume. Moreover, the experiments can be enriched through exploring the effect of the primaries area coverage and the joint effect of the primaries area coverage and the peak luminance level. The combined experimental data can be used to build a perceptual color gamut volume, which can be analyzed further against the mathematical 3D color gamut volume. Due to the assumption that the color volume is not equally used while presenting real images, a statistic analysis of the luminance level of images could be implemented into the mathematical 3D color gamut volume model.

- For the diffuse white level, this is an interesting study to expand. The future work can focus on the direction of measuring diffuse white over a wider range of the images, covering more types of images. Moreover, the comprehensive result can be analyzed against more parameters, or even a model predicting the diffuse white estimation in an image. It should be noted that there are many diffuse white concepts used in HDR research, i.e. the diffuse white from color retouch experts. Therefore, the comparison is also important. Moreover, the estimation of diffuse white was only on the center of the image. It would be interesting to explore if that is location-dependent on the same image content.

- The ‘perfect’ black level in this dissertation works well for the chosen images and these type of artificial patterns. The threshold, visual equivalent to ‘perfect’ black, of more types of images can be measured, i.e. more types of images, the higher dynamic range of images. Since all the images used in the current study are symmetrical, the asymmetrical images could be added to the measurement pool.
• The simultaneous dynamic range model in this study is built based on one single observer’s experimental data. Therefore, one future direction is collecting more average observers/naive observers data to generalize the simultaneous dynamic range model. Furthermore, the simultaneous dynamic range of the higher peak luminance levels should be measured and modeled. The spatial frequency effect on the simultaneous dynamic range was only measured for one stimulus size of two different peak luminance levels. Additional peak luminance levels and stimulus sizes should be measured as well to build a comprehensive model on the spatial frequency effect on the simultaneous dynamic range. Also, applying the experimental data derived model to HDR image rendering would be another direction of the work, especially for some special HDR images.

In addition to the direct future work of these points above, it will be of interesting work about assessing the HDR display’s capability. One part would be of the black level. The model in “perfect” black work showed the requirement of black level to be visually equal to 0 nits/“perfect” black. However, what is the impact of different levels of black on visual experience, and how to characterize such impact with different images? “Perfect” black model is only a starting point about the requirement but not a metric for impact of black level. Also, there will be a difference between presenting images and videos due to different observing time. Therefore, the difference between presenting images and videos would also be worth exploration. Moreover, to build a solid metric to characterize the display’s color reproduction capability, the metric should be depending on the application, even possible of combinations of the statistical usage of the displays. For example, a statistical analysis of the pixel values of enough natural images and rendered images would tell the distribution of the usage of different RGB values of a display, hence a better understanding of which portion of the display volume is used more. Theoretically, the weighted volume should be better correlated with the visual experience. Moreover, one single metric value may not be enough even with the weighted-statistically-based method. It would be benefit to research
on multi-parameter based model for different application. For example, the weighted volume as a general metric with the black level requirement for the dark/dim images and another metric for impact of peak luminance. The psychophysical experiments about perceptual color gamut volume in this study showed some difference from the computational volume. Therefore, maybe a separately stand-alone value for the impact of peak luminance could be of good value in clearly stating the display’s capability considering the peak luminance would have direct impact on the black level requirement.

8.2 Publications


Bibliography


[34] Miller, S., Nezamabadi, M., and Daly, S., “Perceptual signal coding for more efficient usage of bit codes,” SMPTE Motion Imaging Journal 122(4), 52–59 (2013).


166


Appendices
Appendix A

Slice Constant Lightness/Brightness

The new color gamut volume calculation method requires the computation of the maximum chroma at 360 hue angels at constant lightness/brightness. The following code would create the maximum chroma at constant lightness/brightness.

```matlab
function sliceLChJCh( colorAppear, inputfolder, outfolder )

    %UNTITLED7 Summary of this function goes here
    % Detailed explanation goes here
    warning('off');
    if ~isfolder(outfolder)
        mkdir(outfolder);
    end
    matList = dir(fullfile(inputfolder '/*.mat'));
    for mm = 1:length(matList)
        data = load(fullfile(inputfolder matList(mm).name));
        switch colorAppear
            case 'LCh'
                Lab = data.Lab;
                [theta, rho] = cart2pol(Lab(:,2), Lab(:,3));
```

172
theCh = [Lab(:,1), rho, rem(theta+2*pi, 2*pi)/pi*180];
clear('Lab');

case 'JCh'
    theCh = data.JCh;
end

case 'QMh'
    theCh = data.QMh;
end

otherwise
    error('only LCh or JCh');
end

%% get the range
minL = min(theCh(:,1));
maxL = max(theCh(:,1));
lowBound = ceil(minL + 0.5) + 0.5;
upBound = floor(maxL - 0.5) + 0.5;

sliceL = 0.5:1:upBound;
h = 0.5:1:359.5;
chromaHue = zeros(length(sliceL), length(h));

% [huegrid, Lgrid] = meshgrid(h, lowBound:1:upBound);
interpC = polarCinterp(theCh, lowBound:1:upBound, h);
chromaHue(lowBound+0.5: upBound+0.5, :) = reshape(interpC, [], ... length(h));

if lowBound> 0.5
    for hh = 1:length(h)
        singleHue = h(hh);
        L = [0, lowBound:1:(lowBound+10)];
        C = [0; chromaHue( lowBound+ 0.5: lowBound+10.5, hh)]';
        if sum(isnan(C)) *0
            chromaHue(1:(lowBound-0.5), singleHue+0.5) = 0;
        else
            % continue
        end
    end
end
interpC = interp1(L, C, 0.5:1:(lowBound-1), 'spline');
chromaHue(1:(lowBound-0.5), singleHue+0.5) = ... 
    reshape(interpC, [], 1);
end
end

%%% make some adjustment
chromaHue = adjustChrome(chromaHue);
while sum(isnan(chromaHue(:))) | sum(chromaHue(:)<= 0.00001)
    chromaHue = adjustChrome(chromaHue);
end

%%% 
fprintf('finishing: %s\n',matList(mm).name);
save([outfolder matList(mm).name], 'chromaHue', 'sliceL');
end

function out = adjustChrome(in)
    out= zeros(size(in));
    kernelSize = 3;
    %%% add the padding data
    temp = [in(:, (end–kernelSize+1):end), in, in(:, 1:kernelSize)];
    in2 = [zeros(1, size(temp,2)); temp; zeros(1, size(temp,2))];
    for row = 1:size(in,1)
        for col = 1:size(in,2)
            dataPoint = in(row, col);
            if isnan(dataPoint) | dataPoint<0.00001
                out(row, col) = meanWindow( ... 
                    in2(max(1, ... 
                        row-kernelSize+1):min(row+kernelSize+1, ... 
                        size(in2,1)), ... 
                    col:min(col+kernelSize+kernelSize, size(in2,2)))...
The func 'sliceLChJCh' would require the function of interpolation maximum chroma at different hues, as in the following code.

```matlab
function C = polarCinterp( LCh, L, hue)

% polar coordinate interpolation with
%
% LCh is the data used for interpolations.

C = zeros(length(L), length(hue));

hueIntv = 2;

for hh = 1:length(hue)
    Hue = hue(hh);
    if Hue < hueIntv
        idx = LCh(:, 3) <= (Hue + hueIntv) | LCh(:, 3) >= (Hue-hueIntv+360);
    elseif Hue >= (360-hueIntv)
        idx = LCh(:, 3) <= (Hue + hueIntv -360) | LCh(:, 3) >= (Hue-hueIntv);
    else
        idx = LCh(:, 3) <= (Hue + hueIntv) & LCh(:, 3) >= (Hue-hueIntv);
    ```
subLCh = LCh(idx, :);
for ll = 1:length(L)
    interv = 0.1;
    while 1
        idx2 = (subLCh(:, 1) ≥ L(ll) & subLCh(:, 1) ≤ (L(ll) + interv));
        if sum(idx2) ≤ 3
            interv = interv + 0.1;
            if interv > 2.5
                break;
            end
        else
            break;
        end
    end
    targetLCh = subLCh(idx2,:);
    C1 = getmeanC(targetLCh, L(ll), Hue);
    interv = 0.1;
    while 1
        idx2 = (subLCh(:, 1) ≥ L(ll) & subLCh(:, 1) ≤ (L(ll) - interv));
        if sum(idx2) ≤ 3
            interv = interv + 0.1;
            if interv > 2.5
                break;
            end
        else
            break;
        end
    end
targetLCh = subLCh(idx2,:);
C2 = getmeanC(targetLCh, L(ll), Hue);
C(ll, hh) = (C1+C2)/2;
end
end

fprintf('C interpolation compeleted.\n');

function meanC = getmeanC(theLCh, L, h)
    if size(theLCh, 1) <= 3
        meanC = mean(theLCh(:,2));
    else
        dist1 = theLCh(:, 2) .* (pi/180*abs(theLCh(:, 3) - h));
        dist2 = abs(theLCh(:,1) - L);
        dist = sqrt(dist1.^2 + dist2.^2);
        [~, distOrd] = sort(dist);
        meanC = mean(theLCh(distOrd(1:3), 2));
    end
end
end
Appendix B

Look Up Tables for Apple Pro Display

As stated before, the Apple Pro Display’s performance varies with the bright luminance level settings in Figure 7.1. The performance also is stimuli size dependent. Therefore, one fixed size was used in the experiment. Moreover, for the size pattern, the same input digital count would result in different black level, mainly due to the internal reflection. Measurements of the black stimuli were taken for different white luminance levels. These measurements’ result was used to build the LUTs, which was used to generate the digital count of the desired luminance level. The measurements adopted the same as the experimental patterns. The two different patterns were both used in equal amounts, recall left and right examples in Figure 7.1.

Figure B.1 plots the result of the measurements, i.e. the LUTs. The horizontal axis is the request luminance, encoded in digital count. The vertical axis is the actual measured luminance level. It can be found that, the display showed nice performance for request luminance above 1 cd/m². While for request luminance level below 1 cd/m², the actual luminance level is mostly higher than the request luminance level. Moreover, for the same request luminance level, the same input digital count, the actual luminance level is generally increasing monotonically with the peak luminance level except for 252 cd/m² between 0.5 and 1 cd/m². To use the LUTs, use the actual luminance level as the interpolation ramp
data, which will generate the corresponding request luminance/digital count. The LUTs will be used to generate all the experimental images for the simultaneous dynamic range experiments.

Figure B.1: LUTs for different peak luminance levels of the Apple Pro Display.
Appendix C

Code to Optimize the Model for Simultaneous Dynamic Range

For the fitting of the simultaneous dynamic range experiments, the linear fitting is just use "polyfit" in Matlab. For the constrain, all the four lines intersect at the same point, following code can be used to find the intersecting point. This can also be used in finding the intersecting point of the glare-based model.

```matlab
function bestPoint = findAnchorPoint(DataList, x0)
    f = @(x)errorComp(DataList, x);

    options = optimoptions(@fminunc,'MaxFunctionEvaluations',1000, ...
                            'MaxIterations', 1000, 'OptimalityTolerance', 1e-15);

    [x, ¬] = fminunc(f, x0, options);
    bestPoint = x;

function out = errorComp(datalist, x)
    out = 0;
    for nn = 1:length(datalist)
        ...
    end
```
For the nonlinear fitting, similar to the photoreceptor response, the following code can be used to optimize the model parameters. The boundary setting in line 9 can be used to set the constrain, the constant $b$ and $a$ in Eq. 7.4.

```
function [outx, outy] = fitResponseCurve(inputData, x0)
%UNTITLED Summary of this function goes here
% Detailed explanation goes here
f = @(x)errorComp(inputData, x);

options = optimoptions(@fmincon,'MaxFunctionEvaluations',1000, ... 
    'MaxIterations', 1000, 'OptimalityTolerance', 1e-15);

% [x, ∼] = fmincon(f,x0, [],[],[],[],[0 0 0 0],[12 inf inf inf],[], ... 
%     options);
[x, ∼] = fmincon(f,x0, [],[],[],[],[0, 0, 0.2, 0.1],[10, inf, 0.2, ... 
    0.1], [], options);

outx = x;
x1 = x(1);
x2 = x(2);
x3 = x(3);
```
15 \ x4 = x(4); \\
16 \\
17 \text{outy} = x1 \times (1 - \text{inputData}(:,1)^{x2} \div (\text{inputData}(:,1)^{x2} + x3)) + x4; \\
18 \\
19 \textbf{function} \ \text{out} = \text{errorComp(datalist, x)} \\
20 \quad k = x(1); \\
21 \quad n = x(2); \\
22 \quad xo = x(3); \\
23 \quad yo = x(4); \\
24 \quad \text{estY} = k \times (1 - \text{datalist}(:,1)^{n} \div (\text{datalist}(:,1)^{n} + xo) \div yo) \div yo; \\
25 \quad \text{estY} = \log10(\text{estY}); \\
26 \quad \text{out} = \sqrt{\text{mean}((\text{estY} - \log10(\text{datalist}(:,2))).^2)}; \\
27 \textbf{end} \\
28 \textbf{end} \\
29 \\
30 \text{For the fitting of the effect of spatial frequency of the target Gabor pattern on the} \\
31 \text{simultaneous dynamic range, the optimization can be done with the following code, which} \\
32 \text{sets the maximum simultaneous dynamic range between the experimental spatial frequency} \\
33 \text{range of the target Gabor pattern to speed the optimization.} \\
34 \\
35 \textbf{function} \ [\text{outx, dataPoints}] = \text{fittingParabola(inputData, x0)} \\
36 \quad \% \text{this is used in optimization the parabola function of csf} \\
37 \quad \% \quad \text{Detailed explanation goes here} \\
38 \quad \text{inputData} = \log10(\text{inputData}); \\
39 \quad x0 = \log10(x0); \\
40 \quad f = @(x)\text{errorParabola(inputData},x); \\
41 \quad \text{options} = \text{optimoptions(@fmincon,'MaxFunctionEvaluations',1000, \ldots} \\
42 \quad \qquad \text{'MaxIterations'}, 1000, 'OptimalityTolerance', 1e-15); \\
43 \quad [\text{outx, ~}] = \text{fmincon}(f,x0, [],[],[],[],[-\text{inf}, \min(\text{inputData}(1,:),)], \ldots} \\
182
-inf], [inf, max(inputData(1,:)), inf], [], options);

xrange = linspace(min(inputData(1,:)), max(inputData(1,:)), 100);
yrange = outx(1) - ((xrange - outx(2)) ./ outx(3)).^2;
dataPoints = [10.^xrange; 10.^yrange];
outx = 10.^outx;

function out = errorParabola(inputData, para)
    x = inputData(1,:);
    y = inputData(2,:);
    estY = para(1) - ((x-para(2))./para(3)).^2;
    out = sum(abs(estY - y));
end
end
Appendix D

Visual Search with Chroma Series on Different Background Colors
D.1 Introduction

The combination of background and text/figure color has been explored a lot in the past years for better understanding of human visual color system and better human-computer interaction design.

Humans with normal color vision have three types of cone photoreceptor cells: long-wavelength (L), medium-wavelength (M) and short-wavelength (S). Breitmeyer’s study about background on reaction time to stimuli with different size and contrast directly on M cone photoreceptor showed a significant difference between red and green backgrounds, but not much difference between red and blue backgrounds [103]. In addition to that, reaction time was measured along two constant cone axes, L&M and S. The result demonstrated that the reaction time could be derived as a power function of the cone contrast with a baseline [104]. Moreover, the background effect was explored more precisely on upper and lower visual fields through different spot stimuli. The study showed a significant difference between red and green backgrounds, as well as between upper and lower visual fields [105]. Also, a searching-task research experiment showed the difference between foveal vision and peripheral vision [106]. In 1981, Carter reported a study about the reaction time for color differences in CIELUV. The study did not present a mathematical model. Nagy published another study measuring reaction time to targets with different colors from the rest of the stimuli [107]. The study showed a two-stage reaction time to color difference such that a log scale of reaction time is close to a linear relationship with the color difference and becomes constant when the color difference is large enough. However, this study only included four observers and only five different color difference levels. Another study showed evidence from biological measurements of an effect of background on visual activity [108]. All of these studies indicate an impact of background, and contrast between background and target color fundamentally on vision.

In addition to quantitative fundamental research, similar explorations were conducted
to assist human-computer interaction design. Color combinations of background and text/figure/icon, screen type, ambient illuminant, and contrast were evaluated using different methods [109–117]. These studies analyzed the impact in terms of reading speed, reading accuracy, response time, etc. Additionally, studies showed the preference under different color conditions and different types of screens [111–114]. Traditionally, almost all of these studies only used common named colors, i.e. red, green, blue, cyan, etc. Those colors are not well-defined in any color space. For example, the colors used in [118] were just common colors from Windows system software (Powerpoint, Word, etc.). In terms of measuring methodology, a visual search task is the most common approach in these explorations in determining how the combinations impact the observers’ performance. Measuring reaction time or time finishing a searching task is a common critical consideration for researchers. The searching time measurements were not accurate in many previous experiments due to technical limitations. Usually, in measuring searching time participants were all asked to respond as quickly as possible using a push button [106], or keyboard press [103, 104]. There is a time latency between the signal from brain to the movement of the muscles. For the conspicuous target, there is a chance that this latency can be comparable to the time used in finishing the task. Moreover, there were no guarantees that the participants would not “lie” in the experiment by accident or mistake. An eye or gaze tracker device provides a better solution for studying visual search. Human performance on color combinations of character and background was measured by 125 Hz eye tracker and analyzed [118]. Eye movements can also be analyzed with the eye tracker data [118], providing insight into the patterns of a search task. An eye tracker device recording gaze data at a 250Hz sampling rate was used in the present study to ensure an accurate measurement of the search task.

The motivation of this study is to fill the gap between the fundamental cone-excited color target reaction time studies and those using ill-defined color targets. The reaction time to cone-excited color targets is more of a fundamental research work exclusively on visual performance. Those colors are not directly relevant to daily used colors. However, the
reaction time on targets without well-defined colors is not transferable to other platforms or applications. This study aims at measuring the reaction time to a series of stimuli with different color differences against the mean of the rest of the stimuli in a complex image. The chroma series should be aligned on a line in color space, making the chroma series useful in associating with some quantitative variables, i.e. temperature, pressure, speed. For example, gray stands for standard temperature from a sensor in an interface. For real time monitoring, the color would turn to red gradually when the temperature increases and change to blue when the temperature drops. This synchronization would aid in better monitoring the changes, being faster than reading the numbers, especially for complex interfaces. As reaction time is better associated with human’s attention, it is necessary to explore the relationship between reaction time and color difference while not directly mapping color difference onto the variable’s change linearly. Additionally, we compared three different chroma series of hues red, yellow and blue. Furthermore, four different uniform backgrounds were used for measuring the reaction time of the three chroma series. This step was aimed at exploring the optimal choice of background color for an interface design being used mainly for real-time monitoring. Dark and diffuse lighting environments were used to simulate real viewing conditions.

D.2 Experiment Method

The experiment focused on evaluating human visual search and reaction time for finding the most conspicuous colored targets under different lighting conditions and different backgrounds. These will be introduced below.

D.2.1 Apparatus and Setup

The experiment utilized a display, an eye tracker, and two lighting systems in a small laboratory. The display used was a Dell U2415 24” LED-backlight IPS LCD with resolution 1920 ×
Figure D.1: Experiment setup. The bold black arrows are indicating the light directions for spot light and regular office lights. The walls are painted medium gray.
1080, and a maximum white of \((\text{CIEXYZ} = 233.3, 246.9, 255.5)\). The display was calibrated to the sRGB color space \([8]\) using Berns’ and Day’s methods and model \([65, 67]\). The calibration yielded an average error of 0.3 CIEDE2000 color difference unit \([9]\). CIEDE2000 is a standard color difference formula in which one unit of CIEDE2000 is that which humans on average can detect at about 50% chance. It is important to note, this formula was modeled and developed based on normal color vision observers’ data. There is no color difference formula for color vision deficient observers.

The eye tracker used in the experiment was an SMI Red250 mobile, which can record the gaze data at 250 Hz. Figure D.1 illustrates the setup of the experiment. The eye tracker was mounted on the bottom of the display as shown in Figure D.1. This mobile eye tracker does not require a chinrest. The SMI eye tracker allows for a gaze trigger. In this experiment, a trigger was set up in each trial such that if the gaze fixation on the target reaches 1 second, the experiment would automatically proceed to the next step. This would make the experiment more practical and efficient for both observers and analysis.

**Viewing Environment**

This experiment was conducted under three different lighting conditions: dark, ambient lighting, and direct strong spot flare lighting. The room light was turned off for the dark environment. The ambient lighting was provided by fluorescent lights tilted toward the back wall to create a diffuse reflection. The ambient flare reflected from the display was measured in a 3 by 3 grid area. Figure D.2 shows the CIEXYZ values of the 9 blocks. The mean is CIEXYZ = \([11.25, 11.76, 9.956]\) with the standard deviation as \([1.039, 1.089, 0.934]\). The standard deviation is only 0.5% of the peak white of the display (peak white is CIEXYZ = \([233.3, 246.9, 255.5]\)). In the following section, we would just call this the *ambient* condition. The spot light was designed to provide a strong spot flare on the display. However, the data analysis will not be further described because the observers learned the target locations as explained at the beginning of Section D.3.
Target Type and Layout

There were two different types of targets in the experiment: the rectangle and “T” shape. The rectangle was simulating a sensor indicator in a chemical operation interface, and it can also represent a normal variable indicator in any regular interface. The “T” shape has been used historically in vision search tasks [119]. The rectangles were arranged nearly in a grid layout and the “T” shapes were in a strict grid layout.

Figure D.3 illustrates the layouts of rectangle and “T”. Each rectangle and “T” represent one variable in real time system. For both the dark and ambient conditions, randomly one of the red objects was chosen as the target in one test image. This chosen target would be encoded in a specific color. The remaining rectangles/“T”s would be randomly colored in one of the 9 near-neutral colors (see section D.2.1 for details). This design was simulating a real-time system, where the other rectangles/“T”s, representing other variables, would also float in a range. These target positions were picked because they were roughly equidistant to the center of the display, which was the starting position of the search task (see section D.2.2 for details). Figure D.4 shows examples of test images for both target shapes. The blue rectangle in the example on the left is the target, and the yellow “T” is the target in the right image. Additionally, “distractor” images were designed to prevent observers from learning any possible pattern. In each “distractor” image, one of the non-red rectangles/“T”s (see Figure D.3) was randomly encoded with a color. The remaining

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>12.55, 13.15, 11.14</td>
<td>12.06, 12.54, 10.61</td>
<td>12.25, 12.85, 10.83</td>
</tr>
<tr>
<td>11.45, 11.98, 10.15</td>
<td>12.07, 12.55, 10.74</td>
<td>10.63, 11.14, 9.384</td>
</tr>
<tr>
<td>10.00, 10.45, 8.844</td>
<td>10.39, 10.85, 9.226</td>
<td>9.870, 10.34, 8.680</td>
</tr>
</tbody>
</table>

Figure D.2: CIEXYZ values of the measured flare over the 3 by 3 grid over the display.
rectangles would be arranged the same way, randomly colored one of the 9 candidates. Since the “distractor” is no difference from the test image except the target location, examples will not be presented. In total, the experiment consisted of 172 test images and 18 “distractor” images for both viewing environments. Moreover, according to previous research, human vision has better acuity on horizontal and vertical directions over the other directions [46]. In order to compensate for the effect, a second set of test images (both rectangle and “T”) was generated. In the second set, in each test image, a new target location was picked as the closest to 90° counterclockwise to the original target location. Figure D.5 showed the corresponding test images to the examples in Figure D.4. The targets in Figure D.5 are the same colors as those in Figure D.4 with a rotation. However, the colors of the remaining test images are not the same but they follow the same arrangement, randomly assigned from the 9 near-neutral colors. Each observer was randomly assigned one of the two sets for both the dark and ambient environments.

Figure D.3: Rectangle layout (left) and “T” layout (right), the red objects are the potential target positions.

**Background and Target colors**

As mentioned in the motivation, this study was designed to measure and model the reaction time to a chroma series. Three chroma series were picked in this study, chosen along the lines connecting the gray to the display red primary, blue primary and yellow secondary in the CIELAB color space, thus varying in CIELAB chroma. The reason the chroma series were
Figure D.4: Examples of test images for rectangle shape (left) and “T” shape (right).

Figure D.5: Corresponding test images in the rotated set to the examples in Figure D.4.

Figure D.6: Illustration of the three color scales with 1 CIEDE2000 between each consecutive patch.
determined in the CIELAB color space is that the CIELAB color space is relatively more perceptually uniform [66] and the lines from gray to the primaries give the largest range of colors. Figure D.6 shows the three chroma series with 1 CIEDE2000 between any two consecutive color patches. The nine near-neutral colors for non-targets are the first three in the three chroma series in Figure D.6.

Table D.1 lists the CIEXYZ for the rectangle target for both dark and ambient viewing environments. The colors for “T” are different from the rectangle target (utilizing larger CIEDE2000 values) due to the difference between the two different shapes. Table D.2 is the CIEXYZ for “T” shape. These chosen values are based on the pilot test results.

Combinations of four background colors and three chroma series for the targets were explored. The four backgrounds were black, gray, white and the bluish white, which are commonly used in chemical operation interfaces by the Center for Operator Performance (COP). The four colors also have different levels of contrast against the average of these targets, which is neutral gray. The black and white have strong contrast against the target average gray. Bluishwhite has slightly less contrast against the target average. The gray has the smallest contrast against the target average color. The CIEXYZ of the four background colors are shown in Table D.3. Note the difference between the contrast of the background against the target average and the contrast between the targets. Some studies may focus on the contrast between the background and the stimulus only. The main aspect in this study is the contrast between the target stimulus and the average of the rest. Only four different contrast levels between the background and the stimuli were explored in this study.

D.2.2 Task and Procedure

All participants were asked to do both sessions: dark and ambient viewing environments. The eye tracker device was calibrated before each session. The order of the two sessions were randomized for each observer. The order of the images in each session was also randomized for each observer. Figure D.7 shows the schematic overview of the procedure of each session.
Table D.1: CIEXYZ of rectangle target for the dark and ambient viewing environments (Y is in units of cd/m²).

<table>
<thead>
<tr>
<th></th>
<th>X</th>
<th>Y</th>
<th>Z</th>
<th>CIEDE2000</th>
</tr>
</thead>
<tbody>
<tr>
<td>Blue</td>
<td>1</td>
<td>41.97</td>
<td>42.97</td>
<td>51.86</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>42.18</td>
<td>42.64</td>
<td>52.56</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>42.33</td>
<td>42.71</td>
<td>53.34</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>42.24</td>
<td>41.82</td>
<td>55.55</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>42.29</td>
<td>41.42</td>
<td>57.07</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>42.34</td>
<td>41.03</td>
<td>58.62</td>
</tr>
<tr>
<td></td>
<td>7</td>
<td>42.32</td>
<td>40.22</td>
<td>60.95</td>
</tr>
<tr>
<td>Red</td>
<td>1</td>
<td>43.50</td>
<td>44.29</td>
<td>44.74</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>44.34</td>
<td>44.33</td>
<td>43.37</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>44.93</td>
<td>44.64</td>
<td>42.75</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>45.88</td>
<td>44.73</td>
<td>41.46</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>47.17</td>
<td>45.02</td>
<td>39.57</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>47.58</td>
<td>44.84</td>
<td>38.89</td>
</tr>
<tr>
<td>Yellow</td>
<td>1</td>
<td>46.21</td>
<td>49.06</td>
<td>47.55</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>46.71</td>
<td>49.76</td>
<td>46.94</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>47.99</td>
<td>49.76</td>
<td>47.15</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>50.20</td>
<td>53.82</td>
<td>46.79</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>52.50</td>
<td>56.42</td>
<td>46.45</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>54.71</td>
<td>59.53</td>
<td>45.56</td>
</tr>
<tr>
<td></td>
<td>7</td>
<td>57.60</td>
<td>62.53</td>
<td>45.30</td>
</tr>
<tr>
<td></td>
<td>X</td>
<td>Y</td>
<td>Z</td>
<td>CIEDE2000</td>
</tr>
<tr>
<td>---</td>
<td>-----</td>
<td>-----</td>
<td>-----</td>
<td>-----------</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>42.2</td>
<td>42.2</td>
<td>54.0</td>
<td>8.2</td>
</tr>
<tr>
<td>2</td>
<td>42.1</td>
<td>41.3</td>
<td>56.3</td>
<td>10.6</td>
</tr>
<tr>
<td>3</td>
<td>42.3</td>
<td>41.0</td>
<td>58.6</td>
<td>12.2</td>
</tr>
<tr>
<td>4</td>
<td>42.4</td>
<td>39.9</td>
<td>62.6</td>
<td>15.3</td>
</tr>
<tr>
<td>5</td>
<td>42.0</td>
<td>38.1</td>
<td>66.7</td>
<td>18.4</td>
</tr>
<tr>
<td>6</td>
<td>42.2</td>
<td>37.5</td>
<td>70.4</td>
<td>20.1</td>
</tr>
<tr>
<td>7</td>
<td>42.5</td>
<td>35.3</td>
<td>80.4</td>
<td>24.5</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>43.5</td>
<td>44.3</td>
<td>44.7</td>
<td>5.3</td>
</tr>
<tr>
<td>2</td>
<td>44.3</td>
<td>44.3</td>
<td>43.4</td>
<td>7.6</td>
</tr>
<tr>
<td>3</td>
<td>45.9</td>
<td>44.7</td>
<td>41.5</td>
<td>10.6</td>
</tr>
<tr>
<td>4</td>
<td>47.2</td>
<td>45.0</td>
<td>39.6</td>
<td>13.2</td>
</tr>
<tr>
<td>5</td>
<td>48.1</td>
<td>45.2</td>
<td>37.7</td>
<td>15.3</td>
</tr>
<tr>
<td>6</td>
<td>49.7</td>
<td>45.6</td>
<td>35.8</td>
<td>17.6</td>
</tr>
<tr>
<td>6</td>
<td>53.1</td>
<td>45.9</td>
<td>31.8</td>
<td>21.9</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>46.2</td>
<td>49.1</td>
<td>47.5</td>
<td>5.4</td>
</tr>
<tr>
<td>2</td>
<td>48.9</td>
<td>52.2</td>
<td>46.6</td>
<td>8.5</td>
</tr>
<tr>
<td>3</td>
<td>51.4</td>
<td>55.4</td>
<td>46.3</td>
<td>11.0</td>
</tr>
<tr>
<td>4</td>
<td>53.4</td>
<td>57.9</td>
<td>46.0</td>
<td>13.1</td>
</tr>
<tr>
<td>5</td>
<td>56.2</td>
<td>61.3</td>
<td>45.8</td>
<td>15.2</td>
</tr>
<tr>
<td>6</td>
<td>59.0</td>
<td>64.3</td>
<td>44.9</td>
<td>17.4</td>
</tr>
<tr>
<td>7</td>
<td>62.1</td>
<td>68.0</td>
<td>44.8</td>
<td>19.3</td>
</tr>
<tr>
<td>8</td>
<td>65.2</td>
<td>71.9</td>
<td>44.1</td>
<td>21.4</td>
</tr>
<tr>
<td>9</td>
<td>69.4</td>
<td>77.1</td>
<td>43.5</td>
<td>23.8</td>
</tr>
</tbody>
</table>

Table D.2: CIEXYZ of “T” target for the dark and ambient viewing environments (Y is in units of cd/m²).
Table D.3: CIEXYZ of four backgrounds (Y is in units of cd/m$^2$).

<table>
<thead>
<tr>
<th></th>
<th>X</th>
<th>Y</th>
<th>Z</th>
</tr>
</thead>
<tbody>
<tr>
<td>Black (1)</td>
<td>0.23</td>
<td>0.24</td>
<td>0.36</td>
</tr>
<tr>
<td>Midgray (2)</td>
<td>49.04</td>
<td>51.68</td>
<td>56.17</td>
</tr>
<tr>
<td>White (3)</td>
<td>175.45</td>
<td>184.00</td>
<td>201.38</td>
</tr>
<tr>
<td>BluishWhite (4)</td>
<td>174.49</td>
<td>184.43</td>
<td>209.26</td>
</tr>
</tbody>
</table>

Each session started with a black cross superimposed on the center of a full screen gray background. Participants had to fixate on the black cross for 1 second to trigger the test image. In each test image, the observer had to fixate on the correct target rectangle/“T” for 1 second to continue the procedure. If the observer could not find the correct target within 10 seconds, the program would proceed automatically. A gray image would be presented for 3 seconds after each test image to prevent an after-image effect [46]. After the gray image, the black cross showed up again before the next test image. There was no difference between test images and the “distractor” images during the experiment. The only difference is that the “distractor” images were not used for any further analysis. The experiment window closed when the last test image was finished. The observer was instructed to rest after any test image if they felt a need for doing that. The total time to finish each session was around 20-25 minutes. There was a short break between sessions.

Figure D.7: Schematic overview of experiment procedure.
D.2.3 Participants

In total, 26 normal vision observers (age 19-48) and 16 color vision deficient observers (age 19-58) volunteered for the experiment. Among the 16 color vision deficient observers, 14 observers were deuteranopic (missing or weak M cone) and the other 2 observers were protanopic (missing or weak L cone). All observers were tested with the Ishihara Test. The degree of deficiency was not diagnosed since this was beyond the scope of this study. They are all just considered as a group of color vision deficient observers. The average of the color vision deficient observers would be compared with that of the normal color vision observers as a general study. The experiment and recruiting process were approved by RIT’s Human Subjects Research Internal Review Board.

D.3 Result and Analysis

Before the analysis, time used for each trial was extracted from the SMI eye tracker record. The eye tracker kept a record of each event during the experiment, i.e. the starting time of the black cross image, the 1-second fixation finished, etc. The time used in each test image was extracted by calculating the time difference between the starting time of one test image and the time of 1 second fixation or the preset 10-second threshold. Since there is a 1 second fixation time, subtracting 1 second from the time used for each trial would give the actual time used to finish the trial. In total, there are 19628 trials data were collected. There are 40 trials are no larger than 0 second, which makes 0.2% of total trials. These are suspected due to the accumulated 1 second fixation uncertainty. Therefore, these data points were excluded before any processing.

The data are analyzed in terms of two metrics: the mean searching time and the success rate. For a specific threshold time (i.e., 2 second), any trial finished (after a 1-second fixation on the correct target) under the threshold would be considered a success. For the successful trials, two metrics can be derived: the mean searching time of these successful trials and the
proportion of the total trials that were successful, success rate. Both metrics were plotted and analyzed as a function of color difference. In this section, an overall performance of the normal vision observers and the color vision deficient observers were presented and analyzed. The effect of background, the difference between the three chroma series, and the difference between the two different target shapes are then shown.

Again, the spot light flare will not be presented. There were two target candidate positions, close to the center of the spot flare, in the spot light experiment. The “distractors” in the spot light flare experiment can not be determined to be successful in preventing the observers’ awareness of the location pattern according to the fixation data from the eye tracker. Moreover, some of the observers did mention the awareness of the target locations according to the feedback after the experiment. Therefore, the data and analysis of the spot light experiment data will not be presented. Also considering the significant difference between the two shapes (see Section D.3.3 for details), analysis of the rectangle shape data is included in Section D.3.2 to D.3.2. The analysis of the “T” shape and the comparison with the rectangle is presented in Section D.3.3.

D.3.1 Overall Performance

For the overall performance discussion, only the result of the rectangle shape target is used since there is significant difference between the two shapes (see section D.3.3 for more details). The overall result will be shown in two analyses, mean searching time and the success rate. A comparison between the normal vision observers and color vision deficient observers is made.

Mean Searching Time

Figure D.8 plots mean searching time (with standard error) as a function of color difference for 2 second and 4 second thresholds, for both the normal vision observers and the color vision deficient observers. It should be noted that the threshold is set for all trials, which
Figure D.8: Overall mean searching time for 2 second threshold (left) and 4 second threshold (right). The first row is the result of the normal vision observers and the second row is the result of the color vision deficient observers. The horizontal axis is color difference of the target from the mid-gray in unit of CIEDE2000, and the vertical axis is the searching time in second. Red lines are the fitting curves and the markers show mean and standard error.
does not exclude any particular observer data but only the trial data. Therefore, the higher the threshold is, the more trials are included for the mean searching time calculations. The red line in Figure D.8 is the best fitting function in the form of Eq. D.1, where $y$ is the mean searching time, $x$ is the color difference in units of CIEDE2000, and $a$, $b$, $c$ are optimized constant parameters. $a$ represents the gain/curvature; $b$ is the horizontal offset, which was constrained between 3 and 5, because 3 is the maximum color difference of the rest of the target (baseline) and 5 is the maximum such that $y$ will not be negative. $c$ is the asymptotic minimum response time.

$$y = a/(x - b) + c \quad \text{(D.1)}$$

For normal vision observers’ data, the fitting curve works well for both thresholds, with some deviation between the curve and measured data around 10 CIEDE2000. For 5 CIEDE2000, the mean searching time is around 1.5 seconds for 2 second threshold, and it reaches 2.5 seconds for 4 second threshold. The minimal reaction times ($c$) in the two thresholds are very close, $\approx 0.6$ second. One interesting mathematical result is that $(2.5 - 0.6)$ is around twice of $(1.5 - 0.6)$, the same factor as it is for the threshold. For the color vision deficient observers, the measured mean searching time has a larger standard error compared with the normal vision observers’ data for both thresholds. Moreover, the fitting curves do not work as well as for the normal vision observers. For the 2 second threshold, there are three obvious data deviates from the curve. For the 4 second threshold, the data seem split into two trends. Though there was a difference between the two groups of observers’ results, all fitting curves show the $\approx 0.6$ offset. Comparing the best-fit curves, the fitting lines are quite similar for the two second threshold, but a stronger curvature exists for the four second threshold for the color vision deficient observers relative to those with normal vision. The stronger curvature means a similar mean searching time for small color difference (5 CIEDE2000), but a longer mean searching time for large color difference ($\geq 10$
Success Rate

![Success Rate Graphs](image)

Figure D.9: Overall success rate for 2 second threshold (left) and 4 second threshold (right). The red solid line is the best-fitting for the normal vision observers’ data and the blue dash line is the best-fitting for the color vision deficient observers’ data.

\[ y = (x - d)^e + f \]  \hspace{1cm} (D.2)

Success rate (see Figure D.9) is fitted with a power function of Eq. D.2, where \( y \) is success rate, \( x \) is color difference in units of CIEDE2000 and \( d, e, \) and \( f \) are the optimized constants. Here, \( d \) is the color difference offset, \( e \) is the power strength, and \( f \) is the success rate offset. Similar to search time, \( d \) was constrained between 3 and 5.

Figure D.9 shows the success rate with best-fitting curve for both normal vision observers (red line) and the color vision deficient observers (blue line) for a 2 second threshold (left) and 4 second threshold (right). Each data point in the plot is the success rate for the whole group without excluding any data. Since this is just one data point for the group, the standard error is not available for this metric. In general, the function fits well for both the normal vision data and the color vision deficient data. However, clearly it fits better
in the normal vision data and the color vision deficient data split into two trends, which is very similar as in the mean searching time in Figure D.8. On average, the color vision deficient success rate is mostly lower than the normal vision’s, as well as the best fitting line. However, for 5 CIEDE2000, there is no difference between the normal vision and the color vision deficient. When it reaches 10 CIEDE2000, success rate of the normal vision is ≈ 15% higher than that of the color vision deficient for both thresholds.

In general, the two functions fit well with the measured data points. However, the fitting is better for the normal vision observers than the color vision deficient observers. Both metrics showed that the normal vision observers performed better in this searching task than the color vision deficient observers. Another interesting finding from the fitting of the mean searching time is the ≈ 0.6 second offset for all four examples.

D.3.2 Rectangle Layout Result

For a better understanding of the impact from the different variables, analysis of covariance (ANCOVA) was applied. ANCOVA is a tool used for analyzing an independent variable’s effect on a continuous dependent variable. In ANCOVA, a linear regression was fitted between the response and the dependent variable, followed by analysis of the impact of the independent variable on the slope and the intercept. In this analysis, the continuous dependent variable is color difference, a converted version in real calculation, the response is the searching time and the independent variable can be background, viewing condition, observer vision types, etc.

For the rectangle shape layout and "T" shape, the ANCOVA test was conducted separately due to the difference between the two (section D.3.3 for details). For the rectangle shape layout, only the targets with color difference between 6 and 10.5 CIEDE2000 are taken for three reasons. First, the searching time of target with color difference above 10 CIEDE2000 varies very little according to Figure D.8. Second, the very small color difference has very low success rate, which means much fewer responses for the ANCOVA test. Exclud-
ing such extreme small (5 CIEDE2000) would be reasonable. Third, the three chroma series
have similar range and steps between 6 and 10.5 CIEDE2000. Three different thresholds,
2, 4 and 6 second, are used in the ANCOVA test. Before applying the ANCOVA directly,
the color difference is converted using $1/(x - 4.7)$, which makes the converted unit more
linear to the the searching time. This conversion follows the general overall fitting function
in section D.3.1. The details and results of the ANCOVA test are listed in Table D.4. The
table includes four single factors and two interactions. The bold data are those with p value
lower than 0.05, which is the typical significance level. Again, the p value here stands for
the effect of the independent variable, factor in the table, on the covariance between the
searching time and the converted color difference unit.

For the single variable factor, the background showed significant effect for all three dif-
ferent thresholds. The viewing condition, dark and ambient, demonstrated no significant
impact on the observers’ performance. A significant effect from chroma series was found for
2 second threshold, but not for 4 or 6 second threshold. The p value for 2 second threshold
is very close to the 0.05 significance level. Therefore, overall chroma series did not present
a strong impact on the performance. A significant difference was found, however, between
the normal vision observers and the color vision deficient observers.

Two interactions were analyzed as well, chroma series × observer vision and background
× observer vision. Significant impact was found from chroma series × observer vision for
both 2 and 4 second thresholds but not for 6 second. Though there was almost no significant
impact from chroma series alone, the chroma series × observer vision showed a strong impact
on the observers’ performance. Additionally, background × observer vision clearly presented
statistical significant effect on human performance.

**Viewing Environment Effect**

Two different viewing environments were measured in the experiment as described above.
Figure D.10 plots the comparison between dark environment and ambient lighting environ-
<table>
<thead>
<tr>
<th>factor</th>
<th>Threshold</th>
<th>df</th>
<th>mean sq</th>
<th>F</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Viewing Condition</td>
<td>2</td>
<td>1</td>
<td>0.1069</td>
<td>0.55</td>
<td>0.4572</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td></td>
<td>1.66</td>
<td>1.97</td>
<td>0.1602</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td></td>
<td>2.76</td>
<td>1.59</td>
<td>0.2075</td>
</tr>
<tr>
<td>Background</td>
<td>2</td>
<td>3</td>
<td>0.5645</td>
<td>3</td>
<td><strong>0.0296</strong></td>
</tr>
<tr>
<td></td>
<td>4</td>
<td></td>
<td>3.847</td>
<td>4.75</td>
<td><strong>0.0026</strong></td>
</tr>
<tr>
<td></td>
<td>6</td>
<td></td>
<td>9.998</td>
<td>6.02</td>
<td><strong>0.0004</strong></td>
</tr>
<tr>
<td>Chroma Series</td>
<td>2</td>
<td>2</td>
<td>0.60947</td>
<td>3.15</td>
<td><strong>0.0431</strong></td>
</tr>
<tr>
<td></td>
<td>4</td>
<td></td>
<td>0.547</td>
<td>0.65</td>
<td>0.5207</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td></td>
<td>1.305</td>
<td>0.75</td>
<td>0.4708</td>
</tr>
<tr>
<td>Observer Vision</td>
<td>2</td>
<td>1</td>
<td>4.338</td>
<td>22.64</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td></td>
<td>22.058</td>
<td>26.53</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td></td>
<td>29.479</td>
<td>17.12</td>
<td>0</td>
</tr>
<tr>
<td>Chroma Series × Observer Vision</td>
<td>2</td>
<td>5</td>
<td>0.76902</td>
<td>4.06</td>
<td><strong>0.0011</strong></td>
</tr>
<tr>
<td></td>
<td>4</td>
<td></td>
<td>1.9527</td>
<td>2.42</td>
<td><strong>0.0341</strong></td>
</tr>
<tr>
<td></td>
<td>6</td>
<td></td>
<td>1.963</td>
<td>1.17</td>
<td>0.3195</td>
</tr>
<tr>
<td>Background × Observer Vision</td>
<td>2</td>
<td>7</td>
<td>0.7786</td>
<td>4.19</td>
<td><strong>0.0001</strong></td>
</tr>
<tr>
<td></td>
<td>4</td>
<td></td>
<td>4.846</td>
<td>6.05</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td></td>
<td>8.614</td>
<td>5.23</td>
<td>0</td>
</tr>
</tbody>
</table>

Table D.4: ANCOVA analysis result for rectangle shape layout. Bold data are those p values below the significant level, 0.05.
Figure D.10: Comparison of observers’ performance in the dark viewing environment (blue) and ambient viewing environment (red) based on mean searching time (left) and success rate (right) for 2 second threshold.

The ANCOVA test results in Table D.4 also showed statistically close or much higher than the significance level, which agrees with two plots. The ANCOVA result for "T" shape layout in Table D.5 demonstrates the similar conclusion, no significant difference between the two viewing conditions. Therefore, it can be concluded that the 5% diffuse flare does not impact the human searching performance significantly in this task.

**Background Effect**

The ANCOVA test showed significant effect from background on observers’ searching task performance. For a better comparison, Figure D.11 shows the fitting curve to the success rate of the four backgrounds for the normal vision observers (left) and the color vision deficient observers (right) under 2 second threshold. Clearly, both groups of observers showed best
performance, highest success rate, for the mid-gray background while the remaining three are close to each other. The success rate of mid-gray background of the normal vision observers on the left is higher than that of the color vision deficient observer on the right. Overall, the curves of the color vision deficient observers (right) are slightly below those of the normal vision observers.

Figure D.11: Comparison of success rate against color difference among the four backgrounds for the normal vision observers (left) and the color vision deficient observers (right) (rectangle shape only).

**Chroma Series Effect**

Three chroma series were used in the experiment. The ANCOVA showed almost no significant effect from chroma series. Figure D.12 shows the success rate (2 second threshold) as a function of color difference CIEDE2000 for three chroma series for normal vision observers (left) and color vision deficient observers (right). The normal vision observers showed higher success rate on the red series while the color vision deficient observers demonstrated the best on the yellow series. The same can be found for 4 and 6 second thresholds. Moreover, the difference between the three series of the normal vision observers is smaller than those of the color vision deficient observers. This does not conflict with the ANCOVA results as the ANCOVA is applied on the successful trials but the success rate is for the percentage of
successful trials over all tested trials.

\[ \text{Figure D.12: Success rate of different foreground chroma series for 2 second threshold of normal vision observers (left) and color vision deficient observers (right).} \]

D.3.3 Comparison between rectangle and “T”

Target shape was expected to have an impact according to previous studies [109, 120, 121]. Under the same threshold, there was not too much difference in mean searching time because all the successful trials are under the same threshold, which makes it difficult to analyze the target shape difference in terms of mean searching time. However, the other metric, success rate, showed a significant difference for the two shapes.

Figure D.13 includes plots of the success rate of the normal color vision observers at 2 second threshold and 4 second threshold for both rectangle and “T” shapes. It can be seen that the observers showed a much higher success rate for the rectangle than “T”. For 10 CIEDE2000 target, observers performed with a 80% success rate for the rectangle while only a 40% rate for the “T” shape for the 2 second threshold. For 4 second threshold, the success rate reaches 95% for the rectangle and only 70% for the “T” shape. The same difference between the two shapes can be found for the color vision deficient observers as well.

For ANCOVA test of “T” shape, only targets with color differences between 7.5 and 15.5 CIEDE2000 were included (see Table D.2). This constraint was applied for the same reasons.
Figure D.13: Success rate of rectangle and “T” layout for 2 second threshold (left) and 4 second threshold (right).

as for the rectangle shape. Table D.5 lists the ANCOVA test results of four single factors and two interactions, the same as for the rectangle shape layout. For the single factor, viewing condition showed a barely significant effect on observers’ performance, with the p value for the 6 second threshold being close to 0.05. Clearly, background presents significant effect on the performance. Surprisingly, chroma series demonstrated a significant effect, although the p value for 2 second threshold is very close to the significant level. Observer vision type showed a difference from that in the rectangle shape as well. Observer vision has no statistical significant effect on human performance in the “T” shape but clearly significant effect in the rectangle shape. For the interaction of chroma series \( \times \) observer vision, the significant effects of the “T” shape are totally opposite to that of the rectangle shape. For background \( \times \) observer vision, p value for the 4 second is above the significant level while the other two are consistent with those for the rectangle shape.

Overall, background clearly showed significant effect for both layouts. Viewing condition has almost no significant impact on the observers’ performance for both layouts. However, the chroma series and the observer vision demonstrate almost opposite result for the two layouts in terms of significance test. For the chroma series, three steps were chosen for each series for the rectangle while four steps for each series were chosen for the “T” layout. Not
enough steps and two different layouts would contribute to the difference between the two layouts. For observer vision, the significance difference could be caused by such two different layouts. Therefore, the chroma series and observer vision do not have as strong an effect as the background on the observers’ performance. Combining the two layouts, three out of six p values of the chroma series × observer vision are lower than 0.05. However, five of out six p values of the background × observer vision are lower than 0.05. Therefore, background × observer vision clearly has a stronger significant effect than chroma series × observer vision.

D.3.4 Searching Pattern Analysis

With the 250 Hz gaze data from the mobile eye tracker, the analysis of searching patterns is possible. As analyzed before, the two viewing environments data will be combined together since no significant difference was found between the two. In order to analyze the possible searching pattern, the conspicuous test images need to be ruled out. Therefore, only the test images with > 2 second searching time will be included. For the visualization of the gaze data, the raw gaze data from the eye tracker device were extracted with a time line for each observer. For each observer, the eye tracker recorded the gaze in units of the pixels on the display. For a certain time from the beginning of the test image, “one gaze data” was added at the gaze coordinates. Here “one gaze data” is a gaussian distribution centered at the recorded fixation to represent an uncertainty of the location and also for a better visualization. For each accumulated gaze distribution, the eye blinking was excluded. The eye tracker will record (0, 0) as the gaze position when the eye blinked. After excluding the blinking data, the gaze map was normalized by its maximum. It should be noted that all observers were included regardless of color vision status since there is no evidence suggesting that would affect the searching path. Figure D.14 shows the gaze distribution at 0.1, 0.3, 0.5, and 0.7 seconds for the rectangle (first row) and “T” shapes (second row) overlaying the layout plot from Figure D.3. At 0.1 second, an ellipsoid shape distribution on the center is seen. This is suspected to be calibration/fixation uncertainty and the trigger area (which
<table>
<thead>
<tr>
<th>factor</th>
<th>Threshold</th>
<th>df</th>
<th>mean sq</th>
<th>F</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Viewing Condition</strong></td>
<td>2</td>
<td>1</td>
<td>0.0592</td>
<td>0.29</td>
<td>0.5873</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td></td>
<td>0.371</td>
<td>0.38</td>
<td>0.5363</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td></td>
<td>9.788</td>
<td>4.45</td>
<td><strong>0.0349</strong></td>
</tr>
<tr>
<td><strong>Background</strong></td>
<td>2</td>
<td>3</td>
<td>0.7592</td>
<td>4.26</td>
<td><strong>0.0052</strong></td>
</tr>
<tr>
<td></td>
<td>4</td>
<td></td>
<td>2.427</td>
<td>3.01</td>
<td><strong>0.0292</strong></td>
</tr>
<tr>
<td></td>
<td>6</td>
<td></td>
<td>6.356</td>
<td>3.43</td>
<td><strong>0.0165</strong></td>
</tr>
<tr>
<td><strong>Chroma Series</strong></td>
<td>2</td>
<td>2</td>
<td>0.6122</td>
<td>3.06</td>
<td><strong>0.0471</strong></td>
</tr>
<tr>
<td></td>
<td>4</td>
<td></td>
<td>3.9801</td>
<td>4.11</td>
<td><strong>0.0165</strong></td>
</tr>
<tr>
<td></td>
<td>6</td>
<td></td>
<td>15.941</td>
<td>7.28</td>
<td><strong>0.0007</strong></td>
</tr>
<tr>
<td><strong>Observer Vision</strong></td>
<td>2</td>
<td>1</td>
<td>0.3033</td>
<td>1.51</td>
<td>0.2192</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td></td>
<td>0.3196</td>
<td>0.33</td>
<td>0.5657</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td></td>
<td>0.221</td>
<td>0.1</td>
<td>0.7515</td>
</tr>
<tr>
<td><strong>Chroma Series ×</strong></td>
<td>2</td>
<td>5</td>
<td>0.3336</td>
<td>1.67</td>
<td>0.1395</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td></td>
<td>1.7307</td>
<td>1.8</td>
<td>0.1099</td>
</tr>
<tr>
<td><strong>Observer Vision</strong></td>
<td>6</td>
<td>7</td>
<td>5.641</td>
<td>2.59</td>
<td><strong>0.0241</strong></td>
</tr>
<tr>
<td><strong>Background ×</strong></td>
<td>2</td>
<td>7</td>
<td>0.3936</td>
<td>2.21</td>
<td><strong>0.0311</strong></td>
</tr>
<tr>
<td></td>
<td>4</td>
<td></td>
<td>1.475</td>
<td>1.83</td>
<td>0.0775</td>
</tr>
<tr>
<td><strong>Observer Vision</strong></td>
<td>6</td>
<td></td>
<td>4.029</td>
<td>2.17</td>
<td><strong>0.0338</strong></td>
</tr>
</tbody>
</table>

Table D.5: ANCOVA analysis result for “T” shape layout. Bold data are the p values lower than the significance level, 0.05.
is slightly larger than the black cross, first frame in Figure D.7). By 0.3 second, for the rectangle shape the gaze distribution showed a minor cluster to the left rectangle from the center while the majority cluster is still on the center. For “T” shape, most gaze points are at the center with a larger ellipse than that at 0.1 second. Moreover, clearly the gaze showed a bias towards the left in addition to the major ellipse distribution on the center. At 0.5 second, the gaze distribution becomes more distributed for both rectangle and “T”. For rectangles, the distribution shows several clusters at the center rectangle and several rectangles around the center with a clear bias on the top row relative to the bottom row. For “T”, very similar gaze distribution was found, with the distribution of gaze points grouped close to the center “T”s with almost none on the bottom row. By 0.7 second, the gaze distribution showed a very similar distribution to that at 0.5 second but with slightly more onto the last row and peripheral rectangles/“T”s. But the difference is that the largest cluster on the rectangles shifts to the vertical rectangle on the left while the largest cluster is still on the center for the “T”. In general, a search path leaning more to the left and top was found when there is no conspicuous target, and the observers’ gaze started from the center and moved to the periphery.

Figure D.14: Gaze data of all observers at 0.1, 0.3, 0.5, 0.7 second from left to right of the rectangle (first row) and "T" (second row).

In addition to the group average, the gaze data of three random observers are plotted in Figure D.15 separately as colored in red, green and blue of the “T” shape test images.
Figure D.15: Gaze data of random 3 observers colored in red, green and blue at 0.1, 0.3, 0.5, 0.7 and 0.9 second of the “T” shape test images.

Generally, these three observers’ gaze distributions followed the average data, showing a strong bias towards the left of the center at 0.3 second, more on the top-left at 0.5 second, and a more even distribution at 0.7 and 0.9 second with a clear bias at the left-top. Still, difference can be found between the three observers. The green observer showed a bias towards the bottom at 0.3 and 0.5 second, and a more uniform gaze distribution at 0.7 and 0.9 second. The red and blue observers showed similar, close-to-average gaze distributions.

D.4 Discussion

In this study, visual searching on chroma series of three different hues was explored on four different backgrounds, in two viewing environments. Generally, the mean searching time and success rate were analyzed as a function of color difference quantitatively for a threshold time. A modified power function and a modified gamma function were used to fit the mean searching time and success rate, respectively. For the general analysis, the modified power function and modified gamma function fit the data well. The mean searching time shared similarities with the gamma function from previous research data on reaction time measured along two cone axes [104]. Also, the mean searching time agrees well with the two-stage result from Nagy [107], a fast decrease for smaller color differences and a more stable response for large color differences. This study included more observers and a more comprehensive study about the reaction time to color differences for different chroma series and backgrounds compared with Nagy’s work [107]. Moreover, the more developed color difference formula is used in this study.
An interesting finding is that for 2 second and 4 second thresholds the fitting functions showed an $\approx 0.6$ second offset in the $1/x$ function. This is suspected to be the minimal reaction time in this task. This is longer than the traditional reported saccade (150 ms) or express saccade (100 ms) time [122]. It is reasonable that our minimal searching time is longer than a simple saccade time as the search task is more complex. Studies showed increasing searching time with increasing stimuli numbers in an interface [121]. In one complex visual search study, mean reaction time could reach $355 \pm 190$ ms [123] if the observers made a wrong initial saccade. Our searching task image included 20 stimuli, distributed over the whole screen. This is a more complex searching task than the study in [123]. Figure D.16 shows the histogram of search time within 1 second for the combined viewing environments. It shows that the peak of the histogram is 0.5-0.6 second. In this search task, the interface is so complex that making a correct initial saccade movement is unlikely. The complexity of this task makes the mean searching time slightly higher than $355 \pm 190$ ms, which is reasonable and acceptable. The result from this study about minimal search time does not conflict with previous saccade or reaction time studies. Looking forward, search time might be useful as a metric for evaluating task complexity or interface complexity. This minimal reaction time also agrees with the second-stage, constant reaction time for large color differences, from Nagy’s work [107]. The exact minimal reaction time is different in this study from that in Nagy’s study since two different interfaces were used.

Another finding about the overall performance is the discrete data split in the mean searching time in Figure D.8. Since the CIEDE2000 formula is based on the normal vision observers’ data, the three chroma series may not correspond with the sensitivity of the color vision deficient observers. Figure D.17 plots the mean with one standard error of the three chroma series separately for 2 second threshold on the left and 4 second threshold on the right. For both plots, the yellow series and blue series substantially overlap with each other while the red series is clearly above the other two. That means the color vision deficient observers lost the sensitivity on the red series while demonstrating similar response on the
Figure D.16: Histogram of reaction time within 1 second for the dark and ambient viewing environments.

Figure D.17: Mean searching time of the three chroma series for 2 second threshold (left) and 4 second threshold (right). (color vision deficient observers, rectangle shape)
blue and yellow ones. The CIEDE2000 may still predict the color vision deficient observers’ sensitivity well on the yellow and blue but fail on the red. Moreover, most of the discrete data points of the red series have a larger standard error than the yellow and blue. The split of the success rate data in Figure D.9 is also caused by the failing on the red series. The plot will not be presented to avoid repetition. The chroma series separation shows clearly that the red series has a lower success rate than the blue and yellow, which has very similar success rate to the normal vision observers. Analysis of both metric supports the failing of applying CIEDE2000 for the color vision deficient observers. This can not be interpreted as a general conclusion for all color vision deficient observers since the color vision deficient observers in this study are mainly deuteranopic (missing or weak L cone), with 2 protanopic (missing or weak L cone), and no tritanopic (missing or weak S cone). This conclusion could be considered as an average result for the deuteranopic observers.

Before the discussion of the several tested and analyzed variables, 5% flare, observer vision, background, chroma series, etc, it is necessary to emphasize several points of the ANCOVA test in this work. First, the ANCOVA test was only applied on the successful trials under a certain threshold. Therefore, the impact of the variable could be undermined. The success rate should be taken into consideration as well, although the ANCOVA may not show a significant effect for all different thresholds. Moreover, the ANCOVA’s result stands for the significance test among the observer group instead of any individual observer. Due to the limited data points of individual observers under a given threshold, it is difficult to perform the test on individual observers. Therefore, the results are more general. Another interesting point comes in when the p values with different thresholds are compared. Increasing the threshold in the ANCOVA test would add more data points for the test. When the variable itself has a clear significant impact over the observers’ performance, the p value would decrease according to the law of large numbers. However, the variance among the group of observers would also affect the p value. Therefore, the three p values of the same source variable do not necessary decrease, especially when the source variable is not a very
strong factor. For example, all three p values from the background of the “T” layout in Table D.5 are well below 0.05; however, they do not decrease monotonically with increasing thresholds. The three p values of the chroma series in the same Table D.5 decrease with increasing thresholds. The three p values of chroma series × observer vision even increases with increasing thresholds. Therefore, the non-monotonically-decreasing p values with increasing thresholds does not conflict with the law of large numbers.

Comparison between dark environment and 5% of peak luminance ambient viewing environment showed almost no difference on both mean searching time and success rate metrics. Only an ≈ 0.05 second shift (Figure D.10) is found in the fitting curve in mean searching time. Also, for the same color difference, the mean searching time from two lighting environments are within one standard error of each other with only two exceptions. Further, only a 3% success rate difference resulted between the two fitting curves. The ANCOVA test result supported the finding of no significant difference between the two viewing environments. It is suspected that such a low level of ambient lighting, 5% of the peak luminance, does not slow the observers’ reaction time on average as the luminance levels of the chroma series are much higher than this ambient flare. Therefore, the effect from the flare is so small that it is overwhelmed by averaging the observers’ data. As for the impact of a higher level of ambient lighting, it would be worth exploring in a study about the lighting condition for application of such a chroma series, which could be a direction for future work.

Background is the single factor showing a strong impact on the searching task. First, both the normal vision observers and the color vision deficient observers showed highest success rate on the gray background relative to the other three. The ANCOVA test verified the significance impact from the background. This significance test result agrees with the success rate plot. The mid-gray background shows both high success rate and significantly better performance for those successful trials. This outstanding performance of the mid-gray background can be explained by the crispening effect, in which the perceptual contrast/difference between a pair becomes higher when the background has a lower contrast to the average of
the pair [124, 125]. In this study, the pair can be considered as the rectangles/“T”s. Colors of most rectangles/“T” are close to mid-gray. Therefore, according to the crispening effect, the perceptual contrast between those rectangles/“T” becomes bigger over the mid-gray background though the same color difference between the rectangles/“T”s. Several models have been proposed to predict the crispening effect [126–129]. These models predict the crispening effect in terms of thresholds. However, this is not enough to be incorporated into the color difference formula due to those studies focused on achromatic and limited backgrounds only. Therefore, the compensation can not be made for different background at this time. Though not enough for a mathematical correction, these models clearly showed that effect is stronger when the contrast between the pair is smaller. This would explain the difference between success rate curves’ difference, where difference between backgrounds decreases when the color difference increases (see Figure D.11). When the target has sufficient color difference against the non-targets, the observer would quickly find the target, i.e. 100% success rate with the minimal average searching time, which has been discussed above. This finding in this study suggested the existence of crispening effect in not only the static image, where most crispening effect data were collected, but also in this searching task environment. For a practical usage of this study, Table D.6 lists the parameters for Eq. D.2 to predict success rate as a function of color difference for gray background (gray rows) and the average of the other three (plain rows) for the rectangle shape for normal vision observers. This table can be used in aiding human computer interface design when reaction time is critical or for associating the color difference with a real-time changing variable.

The chroma series did not show strong impact over the observers’ performance for rectangle layout but clearly significant effect for the “T” layout according to the ANCOVA test. When comparing different chroma series, only three valid steps were used in the rectangle layout but four in “T” layout. This could be a factor contributing to this opposite significance test results of the two layouts. Moreover, the chroma series could be weak factor in such task, which can affected by stimuli layouts, steps, etc. But the success rate plots clearly
show a difference between the two layouts. Normal vision observers performed better on the red series while the color vision deficient observers performed better on the yellow series. This is not surprising because the color difference formula is only based on normal vision observers’ data. However, the ANCOVA showed again opposite conclusions for significance tests for observer vision. Observer vision is found to be a significant factor for the “T” while not for the rectangle. This is showing that the observer vision type could be a weak factor in terms of significant effect on the task performance. For the two interactions, background $\times$ observer vision (5/6) showed slightly stronger effect than the chroma series $\times$ observer vision (3/6). However, neither of the interactions showed all significant effects over the six ANCOVA tests. Based on this study, clearly, mid-gray background is the best choice for all observers. The best choice of the chroma series is likely to be different for the normal vision observers and the color vision deficient observers, though it is possible that an orange series may perform well for both. More experiments are required to determine the best choice for each of the two types of observers as well as for one series that would work well for all observers.

<table>
<thead>
<tr>
<th>Threshold (second)</th>
<th>a</th>
<th>b</th>
<th>c</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>4.9349</td>
<td>0.1526</td>
<td>-0.4148</td>
</tr>
<tr>
<td></td>
<td>4.2285</td>
<td>0.2460</td>
<td>-0.7826</td>
</tr>
<tr>
<td>3</td>
<td>4.9840</td>
<td>0.0852</td>
<td>-0.2042</td>
</tr>
<tr>
<td></td>
<td>4.7533</td>
<td>0.1878</td>
<td>-0.5169</td>
</tr>
<tr>
<td>4</td>
<td>4.9934</td>
<td>0.0571</td>
<td>-0.1259</td>
</tr>
<tr>
<td></td>
<td>4.8696</td>
<td>0.1403</td>
<td>-0.3515</td>
</tr>
<tr>
<td>6</td>
<td>5.0000</td>
<td>0.0188</td>
<td>-0.0378</td>
</tr>
<tr>
<td></td>
<td>4.9672</td>
<td>0.0728</td>
<td>-0.1642</td>
</tr>
</tbody>
</table>

Table D.6: Parameters for background effects (shaded for gray background, plain white for the average of the other three).
Comparison between the rectangle and “T” shape demonstrated a clear difference between the two. Observers showed a much higher success rate for the rectangle than for the “T”. It is suspected that both target shape and target area contribute to the success rate difference between the two types of targets. One is that the rectangle has a larger size than the “T” shape; rectangle is 9.3 times the area of the “T”. Studies from several decades ago explored the impact of different sizes of letters in search tasks [120, 121]. They concluded that search time decreases with increasing size. The second factor would be the shape. The rectangle is more condensed while the “T” is more extended. One study showed different performance between different icons in a search task [109, 117]. To the best of our knowledge, no systematic quantitative study was found exploring the impact of a shape’s compactness on visual search task. In geography, the Polsby-Popper test is a way to calculate the compactness measure of a shape [130]. Future work could systematically explore the impact of the compactness of a shape. Additionally, no joint effect of shape and size is reported in previous studies. Future work could also involve further comprehensive research on this topic using the shapes with different sizes and different compactness.

To better present the effect of different factors on the observers’ performance, the probability of false rejection is calculated for each factor, including four single factors and two interactions. The probability of false rejection is simply calculated as \( (1 - \prod_{i=1}^{n} (1 - p_i)) \), where \( p_i \) is the p value of each threshold. Table D.7 lists the results, where the overall stands for both layouts. For rectangle and “T”, \( n \) is equal to 3 while for overall \( n \) is equal to 6. If 0.05 is used as the significance level, for rectangle layout the background, observer vision and background \( \times \) observer vision all have a significant impact over the performance. But for “T” layout, none of these variables has a significant impact on observers’ performance. In general, statistically none of these factors/variables has significant impact, although background has the lowest probability of false rejection closest to 0.05. Again, these tests are based on group data instead of any individual observer.

With the help of the high frequency eye tracker, gaze data can be analyzed to explore
any possible patterns to assist human-computer interaction design. For “T” shape layouts, the gaze distribution of the non-conspicuous trials showed a clear group search path bias towards the left at 0.3 second and towards the top row at 0.5 second even until 0.7 second. Two of the three random observers showed almost the same pattern as the average-observer. However, the remaining individual’s search path started towards the left but then moved to the bottom row rather than the top row. For this observer, gaze distribution became evenly spaced around 0.7 second. Therefore, in general a search path starting at the center, moving towards the left and then to the top is expected, with some exceptions. This search path actually coincides with our reading habits mostly starting from top-left. Also, the two shape layouts are both regular. Therefore, it is suspected that the reading habits would unconsciously affect the search path when no conspicuous target is in sight. However, the difference is that the reading habit starts from the very top-left but the search path showed a trend towards the left, top. Therefore, the reading habit may not explain the formation of such pattern entirely. For the rectangle, the gaze distribution (the first row in Figure D.14) has a bias towards the left at the beginning, but not towards the top as strongly as the “T”. The gaze distribution showed a secondary cluster at the left and slightly towards the bottom. Though, both showed some consistency on the very beginning of the search path but then would vary with the exact layout. But, in general, the gaze distribution showed much less

<table>
<thead>
<tr>
<th>factor</th>
<th>rectangle</th>
<th>“T”</th>
<th>overall</th>
</tr>
</thead>
<tbody>
<tr>
<td>Viewing Condition</td>
<td>0.6401</td>
<td>0.8153</td>
<td>0.9335</td>
</tr>
<tr>
<td>Background</td>
<td>0.0325</td>
<td>0.0502</td>
<td>0.0811</td>
</tr>
<tr>
<td>Chroma Series</td>
<td>0.7573</td>
<td>0.0635</td>
<td>0.7727</td>
</tr>
<tr>
<td>Observer Vision</td>
<td>0</td>
<td>0.9157</td>
<td>0.9157</td>
</tr>
<tr>
<td>Chroma Series × Observer Vision</td>
<td>0.3434</td>
<td>0.2525</td>
<td>0.5092</td>
</tr>
<tr>
<td>Background × Observer Vision</td>
<td>0.0001</td>
<td>0.1364</td>
<td>0.1365</td>
</tr>
</tbody>
</table>

Table D.7: Probability of false rejection for two different layouts and the joint of two layouts.)
time spent on the last row, the very left column and the very right column of the rectangles and “T” layouts.

D.5 Conclusions

Experiments were designed and conducted to explore observers’ reaction time over three chroma series under four different backgrounds with two viewing environments. Among the four background colors, observers showed significantly better performance for the gray background, which may be explained by the crispening effect. Among the remaining three background colors, no significant differences were found. The three chroma series, of red, yellow, and blue hues, are found to be different for the normal vision observers and the color vision deficient observers. The normal vision observers showed best performance on the red series, while the color vision deficient observers did best in the yellow series. But the ANCOVA showed different significance test results for the two layouts. Therefore, chroma series is not a strong factor in this study. A search path with a starting bias towards the left from the center was found, followed by a bias to the top for both the rectangle and “T”. This shares some similarity with our reading habits but is not entirely the same. However, this could be taken advantage of by some designers when considering an interface layout. To sum up, a gray background would be recommended for a chroma series that has a starting point at gray. Between the chroma series, the red series would be best for the normal vision observers and a yellow series for the color vision deficient observers. Otherwise, a new chroma series needs exploration. An orange series, would likely work well for both the normal vision observers and the color vision deficient observers. The conclusion can be applied to all displays and all interfaces as long as the display is calibrated well. Moreover, many future directions warrant further exploration from this study, i.e. the search time as a function of target size, target shape complexity, and target compactness as well as additional chroma series.
D.6 Acknowledgements

This work was funded by the Center for Operator Performance (COP). We would also express our appreciation to Dr. Jeff Pelz of RIT’s Multidisciplinary Vision Research Laboratory for providing the eye tracker device.