Implementation of Colour Design Tools Using the OSA Uniform Colour System

by

Jim W. Lai

A thesis
presented to the University of Waterloo
in fulfilment of the
thesis requirement for the degree of
Master of Mathematics
in
Computer Science

Waterloo, Ontario, Canada, 1991

I hereby declare that I am the sole author of this thesis.

I authorize the University of Waterloo to lend this thesis to other institutions or individuals for the purpose of scholarly research.

I further authorize the University of Waterloo to reproduce this thesis by photocopying or by other means, in total or in part, at the request of other institutions or individuals for the purpose of scholarly research.

The University of Waterloo requires the signatures of all persons using or photocopying this thesis. Please sign below, and give address and date.

Abstract

The introduction of colour to the computing environment allows additional information to be provided to the user. There are nontrivial problems which can occur when users are allowed to select the colours to be used. Current approaches to this problem have major drawbacks. There exists in the art world a body of knowledge composed of semi-algorithmic methods for choosing a set of colours, a colour scheme, that is both harmonious and pleasing. This knowledge has yet to be significantly incorporated in digital tools for colour selection. A prototype application to explore the problem of selecting a colour scheme that is both functional and aesthetically pleasing to the user was designed and implemented.

An error-tolerant calibration model applied to the colour CRT monitor allows calibrated colour to be presented with moderate precision on slightly miscalibrated monitors. Simulation of calibrated colour allows the effective presentation of colour spaces with a colorimetric basis. It is hypothesized that colour inaccuracies due to incorrect calibration are similar to those caused by changes in illumination in real world situations, and can thus are compensated for by the human visual system.

The OSA Colour System is proposed as a basis for the selection of harmonious colour schemes, as its structure is amenable to the exploration of colour relationships and colour harmony.

Group theory is used in the creation of a simple and intuitive interface for the prototype application, OSA PlaneSight. The merits of its further application in the construction of user interfaces are also explored.

Acknowledgements

My gratitude goes to Bill Cowan for supervising this thesis and offering support and advice. I also thank my readers, John Beatty, Derick Wood, and Hans-Peter Seidel, for their comments and advice.

Financial support was provided by the National Science and Engineering Research Council of Canada, as well as ITRC/IRC and the Computer Graphics Laboratory of the University of Waterloo. Hardware was provided by Apple Corporation and Digital Electronic Corporation.

I would also like to thank the miscellany of people who helped me in my research, among them the members of the Computer Graphics Lab and the departmental secretaries, for being there when I needed assistance. A fond thanks to the number three, without which life would have been more difficult.

I dedicate this thesis to my father, my mother, and my brother, who set me on the right path.

Trademarks

X11 is a registered trademark of Massachussetts Institute of Technology. Macintosh is a registered trademark of Apple Computer, Inc. PowerPoint is a trademark of Microsoft Corporation.

All other products mentioned in this thesis are trademarks of their respective companies. The use of general descriptive names, trade names, trademarks, and registered trademarks, in this publication, even if the former are not identified, is not to be taken to mean that such names may be used freely by anyone. Nor should unintentional omissions or inaccuracies in this section be taken to mean that certain names are not trade names, trademarks, or registered trademarks.

Contents

1 Introduction **1**

1.1 Motivation 1

1.2 Overview 3

2 Colour and the Interface **6**

2.1 The Perception of Colour 6

2.2 Colour Usage: Functionality *vs.* Aesthetics? 9

2.3 A Primer on Colour Harmony 11

2.4 Colour Selection for the Casual User 13

3 Reflective Surface Models **19**

3.1 Introduction 19

3.2 Light and Surfaces 19

3.3 Grassmann's Laws 21

3.4 On Tristimulus Values, Colour Matches, and Colour Difference . . . 23

3.5 Uniform Colour Spaces 29

3.6 Colour Constancy and Surface Perception 33

3.7 Reflective Colour Systems 35

3.8 The OSA Uniform Colour System 37

4 Calibrated Colour on the Workstation CRT 46

4.1 Introduction 46

4.2 The Shadowmask CRT 47

4.3 Colour CRT Calibration 49

4.4 Gamma Correction: a Model-dependent Calibration Method 52

4.5 Calibration Simulation and Miscalibration Effects 56

5 Group Theory and Interface Design 92

5.1 Introduction 92

5.2 The Macintosh User Interface Guidelines 93

5.3 The Interaction Paradigm 94

5.4 Group Theory in a Nutshell 97

5.5 Group Theory and the Interface 98

5.6 Initial Implementation of the Interface 102

5.7 Casual User Response 104

6 Conclusion 108

6.1 Evaluation 108

6.2 Open Questions 110

A Implementation Details **112**

A.1 The Macintosh Environment 112

A.2 OSA PlaneSight 115

Bibliography **120**

List of Figures

2.1 Twelve colours arranged by hue on a colour wheel. 7

3.1 Schematic diagram of cubo-octahedron, illustrating packed-sphere structure of the OSA lattice. 38

3.2 Cutaway of cubo-octahedron, illustration the orientation of triangular lattice plane $L + g = 0$. 40

3.3 Illustration of plane orientations described by the facets of the cubo-octahedron. 41

4.1 CIE (x, y) chromaticity shifts for OSA plane $L = 0$ when the OSA PlaneSight calibration with $\gamma = 2.3$ is applied to a monitor with $\gamma = 2.2$. 64

4.2 Two views of CIE (L^*, u^*, v^*) shifts for OSA plane $L = 0$ when the OSA PlaneSight calibration with $\gamma = 2.3$ is applied to a monitor with $\gamma = 2.2$. 65

4.3 CIE (x, y) chromaticity shifts for OSA plane $L - j = 0$ when the OSA PlaneSight calibration with $\gamma = 2.3$ is applied to a monitor with $\gamma = 2.2$. 66

4.4 Two views of CIE (L^*, u^*, v^*) shifts for OSA plane $L - j = 0$ when the OSA PlaneSight calibration with $\gamma = 2.3$ is applied to a monitor with $\gamma = 2.2$. 67

4.5 CIE (x, y) chromaticity shifts for OSA plane $L = 0$ when the OSA PlaneSight calibration with $\gamma = 2.3$ is applied to a monitor with $\gamma = 2.8$. 68

4.6 Two views of CIE (L^*, u^*, v^*) shifts for OSA plane $L = 0$ when the OSA PlaneSight calibration with $\gamma = 2.3$ is applied to a monitor with $\gamma = 2.8$. 69

4.7 CIE (x, y) chromaticity shifts for OSA plane $L - j = 0$ when the OSA PlaneSight calibration with $\gamma = 2.3$ is applied to a monitor with $\gamma = 2.8$. 70

4.8 Two views of CIE (L^*, u^*, v^*) shifts for OSA plane $L - j = 0$ when the OSA PlaneSight calibration with $\gamma = 2.3$ is applied to a monitor with $\gamma = 2.8$. 71

4.9 CIE (x, y) chromaticity shifts for OSA plane $L = 0$ when the OSA PlaneSight calibration is applied to a monitor with NTSC phosphors. 72

4.10 Two views of CIE (L^*, u^*, v^*) shifts for OSA plane $L = 0$ when the OSA PlaneSight calibration is applied to a monitor with NTSC phosphors. 73

4.11 CIE (x, y) chromaticity shifts for OSA plane $L - j = 0$ when the OSA PlaneSight calibration is applied to a monitor with NTSC phosphors. 74

4.12 Two views of CIE (L^*, u^*, v^*) shifts for OSA plane $L - j = 0$ when the OSA PlaneSight calibration is applied to a monitor with NTSC phosphors. 75

4.13 CIE (x, y) chromaticity shifts for OSA plane $L = 0$ when the OSA PlaneSight calibration is applied to a monitor with Conrac phosphors. 76

4.14 Two views of CIE (L^*, u^*, v^*) shifts for OSA plane $L = 0$ when the OSA PlaneSight calibration is applied to a monitor with Conrac phosphors. 77

4.15 CIE (x, y) chromaticity shifts for OSA plane $L-j = 0$ when the OSA PlaneSight calibration is applied to a monitor with Conrac phosphors. 78

4.16 Two views of CIE (L^*, u^*, v^*) shifts for OSA plane $L - j = 0$ when the OSA PlaneSight calibration is applied to a monitor with Conrac phosphors. 79

4.17 CIE (x, y) chromaticity shifts for OSA plane $L = 0$ when the OSA PlaneSight calibration is applied to a monitor with P22 phosphors. 80

4.18 Two views of CIE (L^*, u^*, v^*) shifts for OSA plane $L = 0$ when the OSA PlaneSight calibration is applied to a monitor with P22 phosphors. 81

4.19 CIE (x, y) chromaticity shifts for OSA plane $L-j = 0$ when the OSA PlaneSight calibration is applied to a monitor with P22 phosphors. 82

4.20 Two views of CIE (L^*, u^*, v^*) shifts for OSA plane $L - j = 0$ when the OSA PlaneSight calibration is applied to a monitor with P22 phosphors. 83

4.21 CIE (x, y) chromaticity shifts for sample reflectances resulting from illuminant change from CIE A to CIE D_{65}. 84

4.22 Two views CIE (L^*, u^*, v^*) shifts for sample reflectances resulting from illuminant change from CIE A to CIE D_{65}. 85

4.23 CIE (x, y) chromaticity shifts for sample reflectances resulting from change from CIE illuminant A to a fluorescent illuminant. 86

4.24 Two views of CIE (L^*, u^*, v^*) shifts for sample reflectances resulting from change from CIE illuminant A to a fluorescent illuminant. . . 87

4.25 CIE (x, y) chromaticity shifts for sample reflectances resulting from change from CIE illuminant D_{65} to a fluorescent illuminant. 88

4.26 Two views of CIE (L^*, u^*, v^*) shifts for sample reflectances resulting from change from CIE illuminant D_{65} to a fluorescent illuminant. . 89

4.27 CIE (x, y) chromaticity shifts for sample reflectances resulting from illuminant change from CIE D_{65} caused by the addition of white light. 90

4.28 Two views of CIE (L^*, u^*, v^*) shifts for sample reflectances resulting from illuminant change from CIE D_{65} caused by the addition of white light. 91

5.1 Viewing axes and vectors of OSA PlaneSight 95

5.2 Schematic layout of OSA PlaneSight interface 102

Chapter 1

Introduction

1.1 Motivation

The goal of this thesis is the creation of a prototype application to investigate better methods of colour scheme selection using the OSA Colour System. The resultant application is called OSA PlaneSight. Group theory is proposed as an aid in the design of sound, compact, effective interfaces and is employed in the design of OSA PlaneSight.

Colour can be employed on computer displays to provide additional information to the user. Two common applications of colour are to locate information rapidly and to relate widely separated pieces of information. Colour can also be used to distinguish possibly ambiguous interfaces by providing distinct viewing contexts. There are nontrivial problems which occur when the user is allowed to determine the choice of colours to be used for such purposes. These problems have only recently begun to be addressed. Though OSA PlaneSight does not attempt to resolve the conflicts encountered in the selection of colour, it can be used as the basis for tools

that do.

Harmonious colour schemes are aesthetically pleasing sets of colours chosen to accommodate the preferences of the user. The problem of selecting a harmonious colour scheme has been well-studied. Centuries of experience with colour design and colour harmony in the art world have produced many useful generalizations about colour scheme selection. Some of these findings have been incorporated in digital tools for the selection of individual colours and of colour schemes, but much remains to be done. Many current colour selection tools follow the paradigm that a colour is to be selected individually by the user without considering placement of the colour in the intended application context.

Colour selection tools can present colours on computer-controlled monitors in terms of well-studied colour systems. Computer controlled monitors allow colorimetrically precise colour rendition when the monitor has been properly calibrated. The highly controlled conditions required for accurate calibration are far too stringent to be applied in general use. By applying an error-tolerant calibration model to the colour monitor, approximately-calibrated colour can be presented effectively on monitors for which the calibration is not previously known. The implementation of error-tolerant calibration presented here takes advantage of colour constancy, which is the ability of the human observer to "correct" for colour inaccuracies in a scene.

Calibration allows the use of colorimetric spaces, uniform colour spaces in particular. The OSA (Optical Society of America) Colour System, a uniform colour space, is used here in an investigation of the problems inherent in implementing a tool for the selection of colour for a colour scheme to be used on a computer-controlled display. There are several advantages to using uniform colour spaces, particularly the OSA Colour System. Most importantly, distances between colours in the space approximate the perceived colour differences in a uniform colour space,

and the arrangement of the samples in the OSA Colour System is amenable to the exploration and selection of a wide variety of harmonious colour schemes.

Group theory is used during the implementation of a compact interface for OSA PlaneSight as a guide for deciding what controls and operations to allow the user. By using group theory to model a user interaction, it is possible to arrive at a group or set of groups that reaches all states in a minimal number of distinct operations. Observations of symmetries in the states and state transitions are encouraged. This is important in interactions which contain large numbers of states. Operators which take advantage of such symmetries tend to be more compact and intuitive than arbitrary operators, allowing the construction of more easily-used interfaces. In addition, interfaces with fewer controls tend to promote faster familiarity and require less screen space for interaction. In illustration, several operations are demonstrated to be easily modelled under group theory.

1.2 Overview

In Chapter 2, colour is informally introduced. Colour, when used properly, can significantly enhance the quality of a graphic user interface. Some basic principles in colour harmony and colour design are outlined. The integration of these principles with the functional requirements of a user interface is an open problem in general. A sampling of current attempts to address the problems of individual colour selection and colour scheme selection is presented.

In Chapter 3, the perception of light and colour is discussed more formally. The properties of reflected light are of particular interest as they are modelled in the calibration model introduced in Chapter 4. Calibration requires an understanding of basic colorimetry, which attempts to solve the problem of specifying perceived

colours and quantifying differences between them. The phenomenon of colour constancy is presented. Uniform colour spaces are discussed with a focus on a uniform colour space defined by the OSA and its merits as a model for the selection of individual colours and of colour schemes.

In Chapter 4, the operating principles of the CRT are described and a brief overview of calibration techniques for the CRT is presented. As the viewing conditions required for calibration are impractical in the normal working environment, the calibration model for a the proposed colour selection tool must be tolerant of error. A model is proposed that attempts to provide error-tolerant, approximate calibrations by a proper choice of a calibration model and a simulation of a reflective colour system. It is hypothesized that monitor colour shifts due to slight miscalibrations are similar in nature to shifts produced by changing the illumination when viewing reflective objects. This would allow such a tool to make use of colour constancy. Some limited quantitative justification is presented, but a thorough study is beyond the scope of this thesis.

Chapter 5 begins by introducing OSA PlaneSight, a prototype tool to explore the problems of colour selection. The basis for an interaction paradigm for OSA PlaneSight is proposed. Group theory is presented as an aid in the design of user interfaces. A simple, intuitive interaction paradigm for OSA PlaneSight is then created, illustrating the use of group theory. The initial interface design is presented along with an informal study of user responses to the interface. Resulting modifications of the interface are then discussed.

In Chapter 6, conclusions are presented and some open problems are discussed.

In the appendix, implementation details of OSA PlaneSight on the Macintosh II personal computer are covered. The essential aspects of the application should be easily ported to other computer systems that provide reasonably free access to

colour resources.

Chapter 2

Colour and the Interface

2.1 The Perception of Colour

Before colour can be discussed in relation to the interface, it must first be defined. Light is a form of radiant energy, and possesses intensity, wavelength, and frequency. The spectral composition of light is described by its intensity over a given spectrum, or range of wavelengths. Colour in the psychophysical sense is the "characteristic of a visible radiant power by that an observer may distinguish differences between two structure-free fields of view of the same size and shape such as may be caused by differences in the spectral composition of the radiant power entering the eye." [WS82]

More simply put, colour is the visual sensation produced by light interacting with the retina of the eye that allows one to distinguish between radiant power distributions that differ over the visible spectrum. Colour is also empirically defined in terms of the spectral emission of radiant power. This definition is covered later in Chapter 3 as it is not immediately relevant to the discussion below.

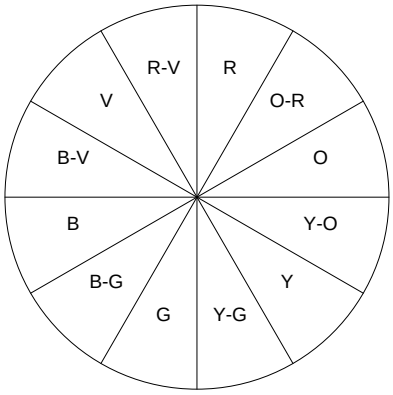

Figure 2.1: Twelve colours arranged by hue on a colour wheel.

Colour schemes are discussed in Section 2.3. Some terminology relevant to this discussion is introduced below.

The colours that most humans perceive are often described in terms of *hue*, *value*, and *chroma*. Hue is a property of chromatic colours that allows them to be classified in terms such as red, yellow, green, or blue, or as a gradation, or intermediate colour, between other colours. A colour is achromatic if it appears as a white, black, or grey. Achromatic colours, also known as neutral colours, are said to have no hue. Value refers to the degree of lightness in a colour; black has minimal value, while white has maximum value. Chroma refers to the degree of saturation of a colour, or its intensity or purity. The chroma of a colour is defined in terms of its difference from a neutral colour of the same lightness. Neutral colours have no chroma. Saturation refers to the perceived purity of a colour, while chroma is a physically measured quantity.

Colours can be arranged in a circle, with hue changing as one moves about its circumference. Artists typically use hue circles, or colour wheels, with twelve reference colours. The hues change gradually as one goes around the circle, passing through red, red-orange, orange, yellow-orange, yellow, yellow-green, green,

blue-green, blue, blue-violet, violet, red-violet, and back to red. This relative arrangement is illustrated in Figure 2.1; the colours red, orange, yellow, green, blue, and violet are denoted by their first letters. The colour wheel was developed in the context of paints to allow the artist a convenient means of arranging colours. The three primary colours, namely red, yellow, and blue, are arranged in an equidistant triangle on the colour wheel. A given primary colour has the property that it cannot be obtained by a mixture of the other two associated primary colours; all colours can be obtained by mixing primary colours in varying amounts. The purest colours are placed on the boundary of the colour wheel. Mixing various hues tend to produce neutral colours, which are placed in the interior of the wheel; totally neutral colours lie at the center. When paints are mixed, the result is darker than the composing pigments. Ideally, paints can be mixed to form grey and black. The results of mixing real paints is somewhat complex, and is generally beyond the scope of this thesis.

The three primary colours for mixing light, instead of pigment, are red, green, and blue. The hues of two lights are *complementary* if they can be added to produce white. The hues of such hue pairs lie opposite each other on the colour wheel. The hues of two pigments are complementary when they can be mixed to produce a neutral colour, not necessarily grey. Two hues are *near-complementary* if they are merely close to being opposite on the colour wheel, but not exactly.

Two colours are *analogous* if they have a hue component in common; this is generally interpreted as meaning that the colours are within 60 to 90 degrees of each other on the colour wheel. Analogous hues can be used to imply similarity.

2.2 Colour Usage: Functionality *vs.* Aesthetics?

Colour is used by people to gain information about their environment, both natural and manmade. Colour is also used by humans to indicate function and to provide information[Cow89a]. For instance, red can be used to indicate a warning or danger; green can indicate that a fruit is not yet ripe or be a signal that one can proceed in traffic. Colour can be used similarly in an interface to convey information to a user. For example, colours can be used in an interface to indicate warnings and error conditions, or to distinguish or relate functionally distinct parts of an interface. The colour context of an interface refers to the information than can be inferred from a given colour usage in the interface.

The increased use of colour in computer interfaces over the past decade has largely been made possible by the reduced cost and increased power of colour graphics display hardware. The large gamut of colours available on modern colour displays raises the difficult problem of judicious selection of colours for use in an application.

For many applications the colours used are chosen by implementation teams consisting of programmers, analysts, and graphic designers. For the most part, such applications do not allow displayed colours to be altered by the casual user. The result may or may not be aesthetically pleasing for a given user since what is considered pleasing varies from person to person and from era to era. There is a general trend towards functional use of colour with little regard to aesthetics. This trend is a reaction to the tendency of untrained colour choices based on personal aesthetics to provide a suboptimal amount of additional information. A poor choice of colour scheme interferes with the functionality of the interface by being distracting or by providing confusing or superfluous information.

The problem increases in complexity on a window-based interactive display when several applications coexist on different parts of the same display. The context of colours employed by one application should not interfere with the contexts of colours used by other applications. A solution to this problem is to provide a facility that manages the sharing of certain designated colours, thus providing a more coherent interface for the user. In the absence of such a facility, colour selection for each application is made without regard to that of other applications. The outcome is an interface that is often confusing to the user because the meaning of colours may vary wildly from one application to another.

A conservative solution to this problem is to severely restrict colour usage on the display. A monotonous choice of colours made in the interest of visual efficiency may be self-defeating, however. Humans are adapted to living in a colourful environment, desire colour in their environment, and respond favorably to it. They do not react well to monotonous environments. When people are faced with monotonous stimuli, their attention spans tend to be impaired, resulting in more frequent errors. This has been widely documented in the literature. For instance, Hockey[Hoc83] noted that "Research on the consequences of exposure to sensory deprivation demonstrated that perceptual and cognitive processes were severely impaired, and led Hebb[Heb55] to emphasize the contribution made by sensory variation to the preservation of the efficient functioning of the brain."

Instead of taking such a conservative approach, it is preferable to have a mechanism that provides a rich set of functional colours to which each application could refer. If the selection of colour in such a mechanism can be automated, it is reasonable to expect the mechanism account for users' preferences in its colour selection strategy, arriving at a compromise colour usage that is both functional and aesthetically pleasing. The creation of such a mechanism requires designers and

programmers to have a thorough understanding of functional colour. There exists such a body of experience and knowledge in the art world upon which can be drawn from when constructing such mechanisms.

2.3 A Primer on Colour Harmony

Colour is sufficiently understood in the field of art that general strategies have been formulated for the selection of aesthetically pleasing colour schemes[Bir88, LSP89, Qui89, Won87]. Some of these basic approaches are discussed below. The lack of a precise mathematical formulation in these strategies reflects the intuitive nature of the artist's understanding of colour and colour harmony. Some terminology must be defined before we can proceed with a discussion of the heuristic and semi-algorithmic strategies used by artists.

A *colour scheme* is a set of colours selected for a design. From a user interface design standpoint, an ideal colour scheme is both functional and aesthetically pleasing. For a colour scheme to be aesthetically pleasing, or harmonious, it must provide a large enough number of colours and a wide enough range of colour to keep the display from appearing monotonous and fatiguing. A good choice of colour scheme should also enhance the functionality of an interface.

Colour schemes can be generated based on *value gradation*, *chroma gradation*, *hue gradation*, *hue mixtures*, *split-complementary hues*, or *unrelated hues*. These schemes are described below.

A colour scheme is termed *monochromatic* if all colours in the scheme share the same hue. *Value gradations* can be used to create a strong appearance of depth. Regions of strong contrast tend to attract the eye.

Colour schemes using *chroma gradation* are most effective when value does not vary greatly. Dark backgrounds heighten the contrast between the colours and make the foreground colours appear more luminous. Chroma gradations are not easily manipulated to give the illusion of depth. Differences in value tend to dominate in a depth illusion, rather than differences in chroma.

In practice, value and chroma cannot both be kept constant while varying hue since hues at full chroma vary significantly in value for both paint and monitors. This is primarily due to variation in the sensitivity of the eye over the spectrum.

For *hue mixture* colour schemes, two colours and their mixtures are used to produce a range of *hue gradation*. If two analogous colours are mixed, the chroma of the mixture remains approximately as strong as the original colours. If complementary colours are picked, some mixtures show a significant loss of chroma.

Hues can also play a role in the generation of spatial illusions. Warm hues, such as red, yellow and orange, are associated with fire. Cool hues, such as blue, violet, and green, are associated with water or the sky. Warm hues appear to advance while cool hues appear to recede. Though hue differences can be used to create spatial illusions, differences in value play a larger role than hue differences in these effects. The human vision system is more sensitive to changes in value than in hue or chroma.

When employing hue mixtures, one must take care to avoid side effects arising from *simultaneous contrast*, the phenomenon of apparent changes in hue, value, and/or chroma of colours when one colour is placed adjacent to another. This phenomenon can strengthen or reduce perceived differences in hue, value, and chroma. Simultaneous contrast is most apparent when one colour is surrounded by another; the surrounded colour appears to change. In this case, the effect is highly dependent on the relative sizes of the enclosed colour and its surrounding colour.

Split-complementary colours schemes are among the most varied and interesting. A split complementary scheme is obtained by replacing one of the two hues in a complementary pair, the *split hue*, with the two hues adjacent to it on the colour circle. Alternately, one can retain the retain the split hue in the scheme, or even split both hues for a larger range of hue. Near-complementary hues can be split similarly.

A scheme employing *complementary hues with hue gradations* is produced by mixing a split hue in a split-complementary scheme with its adjacent hues to produce colour gradations. Combining complementary hues with value gradations can give an illusion of spatial separation between adjacent parts of the design. The use of chroma gradations with complementary hues provides for strong accentuation of the colours of full chroma.

Triads, sets of three hues 120 degrees apart on the colour circle from each other, can be combined in a colour scheme. Colour schemes can be constructed using *unrelated hues* by careful determination of value and chroma, or by the result of intuitive colour selections. Tinting the colours of a scheme using unrelated hues with a common colour can help establish some analogy between dissimilar hues, making the appearance of the scheme more harmonious. Such schemes are termed *common-denominator* colour schemes.

2.4 Colour Selection for the Casual User

There are several mechanisms which address the problems of colour selection by a casual user. These mechanisms vary significantly in sophistication, ease of use, and capability. We will briefly discuss several mechanisms in current use and their shortcomings.

Colour selection refers to one of two different procedures: selection of a single colour, and selection of a colour scheme. Procedures to select single colours are generally part of procedures to select entire colour schemes, and are therefore discussed first.

One of the simplest ways of selecting a single colour is to allow the user to manipulate each of the three monitor phosphor outputs separately — red, green, and blue (RGB) — and then see the resulting colour. A casual user must experiment to learn how additive colour mixing works on a monitor in order to use this scheme effectively.

In the standard Macintosh colour picker utility, monitor RGB values are mapped onto a hue, saturation, and value, or HSV, model. The utility displays a colour circle as a navigational aid, as well as the colour that is being changed and its original colour. A colour is selected by specifying numeric values in monitor RGB or HSV coordinates, or by manipulating the colour circle. The hue and saturation of a colour are selected by specifying a position on the colour circle; the value of the colour and of the colour wheel is determined by a slider control, which allows the user to select a point in a given range. The selection is performed against a white background, since the apparent colour of the selected colour may be different on other backgrounds. Experiments [SCB87] indicate that both the HSV and RGB models are roughly equal in ease of use.

The X Window System, referred to as X, uses an entirely different scheme for the selection of colour. X is a portable network-based window system. In release 4 of version 11 of X, a user can specify the colours to be used in specific parts of an application's display by making text entries in a resource file or by specifying parameters to an application at the start of execution. Other parts of the interface may also have colour specified, depending on the window manager being used.

Colours for such casual use in X are specified by text names, which are mapped onto 24-bit RGB values. There is no way for the casual user to specify additional named colours, as the translation from names to RGB values is maintained in a global system file. There are about 500 distinct named colours available, 101 of which specify monitor grey levels from 0% to 100%. Colour names can be quite useful when referring to colours of real objects such as "forest green" and "salmon". However, the naming of colours can be arbitrary and nonintuitive. For example, what colours are represented by "firebrick" or "peru"? The differences between "burlywood1", "burlywood2", "burlywood3", and "burlywood4" are not obvious, though the increasing number indicates decreasing RGB intensity: the designations "2", "3", and "4" correspond to approximately 93%, 80%, and 55% respectively of the RGB intensity of the "original" colour, designated by "1". The actual monitor intensities produced for such a group of colours do not form an evenly-spaced sequence. The reason for this is the "gamma factor", which is discussed in Chapter 4. The end result of all these problems is that searching for a desired colour by name can be a tedious process of trial and error.

RGB selection, HSV selection, and named-colour selection comprise the three procedures for individual colour selection in common use. The selection of colour schemes is a more complex problem, and thus a greater variety of solutions have been tried.

While the selection of individual colours is relatively well-supported, there is a lack of standard tools to view these individually selected colours in the context they are to be used in. Because of this lack, the process of selecting a colour scheme to fit one's aesthetic preferences often turns into a procedure of sequentially picking a colour and then determining which of the previously selected colours must be changed to accommodate the newly added colour.

In the case of picking one colour, only one colour need be selected. For the case of two or more colours, an iterative algorithm of selecting an nth colour and comparing it to the $n-1$ colours already picked, each of the $n-1$ colours may have to be adjusted to accommodate the new colour, resulting in a maximum of $n-1$ additional colour selections. In the worst case, the number of selections thus required by this algorithm is $O(n^2)$ selections, assuming very optimistically that each selection requires constant time. The actual time required may be much worse, as the nth colour being considered may not be amenable to the colours already picked.

If each colour already selected requires one additional correction to rebalance the colour scheme, the time required increases to $O(n^3)$ time. If earlier colour selections are found to be inappropriate, repeated changes of previous colours may result in an algorithm which is bounded only by the patience of the user and his or her tolerance for the imperfection of the colour scheme being selected. A more realistic bound is probably somewhere between $O(n^3)$ and $O(n^4)$.

This technique, not unique to X, rapidly becomes time consuming as the number of colours increases, particularly when the application which uses desired colours must be stopped and restarted with each colour change. An increase in the number of available colours does not increase the time taken to locate a desired colour proportionately; the method of navigation can greatly affect the time taken for the selection of a desired colour.

Notice that the selection of colours in X as described above is *static*. That is, colour choices are determined and fixed during initialization of an application. Changing the colour choices requires terminating and restarting the application with new colour choices.

Another solution is to allow the user to select a small number of colour prefer-

ences, and attempt to generate automatically a colour scheme that is both functional and aesthetically pleasing based on a few individual colour selections.

PowerPoint, an electronic slide show presentation application by Microsoft that runs on the Macintosh, is a case in point. PowerPoint allows a colour scheme of eight colours to be applied to one or more slides. There is a predefined set of 90 colours from which to choose, with the option to resort to the Macintosh colour picker utility if more flexibility is desired. Individual colours in the scheme can be selected, just as before; in addition, colours can be fine-tuned via the Macintosh colour picker. Colour allocation is *interactive*, meaning that colour choices can be changed during execution of the application.

PowerPoint has an option that allows an entire colour scheme to be selected at one time, greatly reducing the number of colour choices that must be made compared to the trial-and-error technique described earlier. The user is allowed to choose the background colour from the set of 90 stock colours. The user then chooses a foreground colour from a subset of the stock colours, constrained by the program according to the background colour selection. Finally, the user is allowed to choose one of several sets of six colours chosen from the remaining stock colours that presumably provide an aesthetically pleasing colour scheme. This method is not infallible; by picking disharmonious foreground and background colours, the resulting colour scheme may be quite jarring. While the schemes provided may serve the purpose of a slide show by attracting attention, they are likely to be inadequate for extended use in a workstation environment.

In the examples given above, the colour selection mechanisms provided restrict the number of colours available. Many computers now provide eight bits of resolution for each of the three monitor primaries. One obvious alternative would have been to allow full freedom of navigating the 2^{24} colours; navigating and searching

through such a large space of individual colour choices is potentially laborious and time-consuming. By trading off the range of colour in colour selection, the time required to arrive at a individual colour selection is generally reduced. However, a colour range that is too small may not be able to accommodate aesthetic usage and personal taste when selecting a colour scheme.

Navigating a colour space with the goal of selecting a colour is a problem which has not been resolved in general. Power Point, which has only 90 colours from which to pick from, has a relatively simple means of navigating and exploring all the available colours and most of the allowed colour palettes. X on the other hand, with approximately 500 distinct colours, does not provide a simple means of navigating the named colours and has no built-in provisions for selecting colour schemes.

In conclusion, while the problem of selecting a single colour has been well-studied, the problem of selecting a colour scheme on a system capable of a large range of colour has yet to be solved. The most common approach has been to let the user experiment by selecting one colour at a time and altering the colour scheme dynamically. Some tools, by offering a choice of a colour scheme based on a few initial colour selections, such as Power Point, attempt to reduce the number of individual colour selections required to arrive at a colour scheme, but the choices available are generally too restrictive. From the example of Power Point, it is clear that some form of colour choice automation can save a great deal of time in the colour scheme selection procedure. It is also desirable to have a large range of colour choices, which is reflected in the implementation of X. An ideal tool would have both of these strengths.

Chapter 3

Reflective Surface Models

3.1 Introduction

Light and the human perception of colour are discussed more formally in this chapter. The phenomenon of colour constancy is examined. The theoretical and practical bases of several colour models are explored in order to justify the choice of the OSA Colour System. It is hypothesized that OSA PlaneSight can take advantage of colour constancy when using the OSA Colour System and present the appearance of calibrated colour on a computer controlled monitor. Quantitative justification of this is presented in Chapter 4.

3.2 Light and Surfaces

Light, or radiant energy, is composed of photons. Photons have the properties of *energy* (E), *frequency* (v), and *wavelength* (λ). These properties are related by the

following formula:

$$E = v/\lambda. \tag{3.1}$$

Frequency is expressed in Hertz (cycles per second), which are abbreviated as "Hz". Wavelength is often expressed in nanometers, which is abbreviated as "nm" (1 nm $= 10^{-9}$ meter). The visible region of the spectrum is usually taken to be the range from 380nm to 760nm; the sensitivity of the human eye drops to well below 1% of its maximum sensitivity at the ends of this range. The actual region considered varies depending on the intended application, as the normal human eye is capable of detecting frequencies beyond those limits if the radiation is sufficiently intense. Ranges as wide as 360nm to 830nm [WS82] have been proposed for use. The term *monochromatic* describes a spectral distribution composed of photons at a single wavelength.

The colours perceived in the world by the human visual system are mostly produced through absorption and reflection. Surfaces may absorb a given frequency completely or partially, the remainder being reflected. Problems introduced by the phenomenon of *luminescence* are ignored in this discussion. Luminescence is the result of energy being absorbed at given frequencies and then re-radiated at other frequencies.

The light reflected by a surface can be described as a function $\beta(\lambda)$ of frequency. A perfect reflector is given by $\beta(\lambda) = 1$. Given a function $S(\lambda)$ that describes the light incident on a surface, the light immediately reflected is $E(\lambda) = S(\lambda)\beta(\lambda)$. The light incident on a surface or sensor is highly dependent on the geometry of the scene. This light can be measured with sensors; with additional experimental data, the perception of colour that is associated with this light can be quantified.

Colorimetry is concerned with the specification of colour matches and colour

differences. A complete colour match is made if and only if stimuli look alike when viewed under the same viewing conditions by observers with normal colour vision. A quantitative specification for colour matches must predict colour matches accurately. When a colour match is not observed, the usually small colour differences are specified using colour-difference formulae, which are derived using experimental data. A numerical specification of the colour of a physically-defined visual stimulus must be composed of continuous functions of the physical parameters defining the spectral radiant power distribution of the stimulus. The experimental laws governing colour matching are summarized in an empirical generalization known as the *trichromatic generalization*. These laws are discussed at some length in the following section.

3.3 Grassmann's Laws

Grassmann's laws of colour addition[WS82], which are listed below, characterize normal human colour perception. Three empirical laws have been found to hold over a wide range of viewing conditions.

1. The eye can distinguish only three kinds of variation.
2. If, of a three-stimulus mixture, one stimulus is steadily changed while the others are held constant, the colour of the mixture changes steadily.
3. Stimuli of the same colour (measured using the three kinds of difference) produce identical effects in mixtures, regardless of their spectral composition.

These laws are said to be the *qualitative form* of the *trichromatic generalization*. These laws also comprise the weaker trichromancy principle, termed weaker since

they do not assume the results of colour matching obey certain linear relations. The stronger *quantitative form* is obtained by assuming such linearity. "Over a wide range of conditions of observation, many colour stimuli can be matched in colour completely by additive mixtures of three fixed primary stimuli whose radiant power levels have been suitably adjusted"[WS82], either by mixing the three primary stimuli together, or mixing one or two of the primary stimuli with the desired colour stimuli. There is much freedom in the selection of the three primary stimuli, though the choice is not entirely arbitrary in that they must be chosen so that no one of the stimuli can be matched by a mixture of the other two.

The additive mixture of two colour stimuli is defined as the colour stimulus for which the radiant power over any given wavelength interval in any part of the spectrum is equal to the sum of the radiant powers of the two given colour stimuli for that same interval, assuming the latter two are *incoherent*. Two light sources are said to be incoherent when their phase difference varies with time, causing the light at the frequencies in common to reinforce and cancel with each other over time. *Coherent* light is seldom found in nature; lasers are the only practical source of coherent light.

The quantitative form of the trichromatic generalization is also obtained when we assume the following empirical laws, which have been found to hold over a wide range of viewing conditions:

Law of Symmetry

If a colour stimulus **A** matches a colour stimulus **B**, then colour stimulus **B** matches colour stimulus **A**.

Law of Transitivity

If **A** matches **B** and **B** matches **C**, then **A** matches **C**.

Law of Proportionality

If $\mathbf{A}$ matches $\mathbf{B}$, then $\alpha\mathbf{A}$ matches $\alpha\mathbf{B}$, where α is any positive factor by which the radiant power of each of the colour stimuli is increased or reduced, when its relative spectral distribution is kept constant.

Law of Additivity

Let $\oplus$ denote the mixing of two colour stimuli. $(\mathbf{A} \oplus \mathbf{C})$, $(\mathbf{B} \oplus \mathbf{D})$, $(\mathbf{A} \oplus \mathbf{D})$, $(\mathbf{B} \oplus \mathbf{C})$ denote the additive colour mixtures of $\mathbf{A}$ with $\mathbf{C}$, $\mathbf{B}$ with $\mathbf{D}$, $\mathbf{A}$ with $\mathbf{D}$, and $\mathbf{B}$ with $\mathbf{C}$, respectively. For any four colour stimuli $\mathbf{A}$, $\mathbf{B}$, $\mathbf{C}$, $\mathbf{D}$, if any of the two of the following three colour matches holds, namely $\mathbf{A}$ matches $\mathbf{B}$, $\mathbf{C}$ matches $\mathbf{D}$, and $(\mathbf{A} \oplus \mathbf{C})$ matches $(\mathbf{B} \oplus \mathbf{D})$, then the match $(\mathbf{A} \oplus \mathbf{D})$ matches $(\mathbf{B} \oplus \mathbf{C})$ also holds.

This stronger form of the trichromatic generalization as stated above is directly implied by Grassmann's laws of additive colour mixture. This generalization provides a formal definition of colour, and thus of colour systems, in terms of additive mixtures of three primary colours.

3.4 On Tristimulus Values, Colour Matches, and Colour Difference

The construction of an appropriate colour model can make it possible to quantify colour difference and related phenomena. Tristimulus space is the basis of most commonly-used colour models; in tristimulus space, each perceived colour can be measured and uniquely identified by a triad of tristimulus coordinates. These tristimulus coordinates form a three-dimensional colour space, called tristimulus

space; the tristimulus coordinates of a point in such a space are also referred to as tristimulus values.

The trichromatic generalization allows a colour $\mathbf{C}$ to be specified as matching an additive mixture of any three linearly-independent fixed primary stimuli. As we now show, it is easy to convert between tristimulus values relative to one set of primaries $\mathbf{P}_1$, $\mathbf{P}_2$, $\mathbf{P}_3$ and tristimulus values relative to another set of primaries $\mathbf{Q}_1$, $\mathbf{Q}_2$, $\mathbf{Q}_3$. Let $\sim$ denote a colour match and $\oplus$ denote colour summation.

$$\mathbf{C} \sim \bigoplus_{j=1}^{3} p_j \mathbf{P}_j \sim \bigoplus_{i=1}^{3} q_i \mathbf{Q}_i \tag{3.2}$$

where $\sim$ indicates a colour match, the p_j and q_i are constants.

The primaries $\mathbf{P}_j$ can be expressed in terms of a new set of primaries $\mathbf{Q}_i$ as follows. For each $\mathbf{Q}_i$ we have

$$\mathbf{Q}_i \sim \bigoplus_{j=1}^{3} m_{ij} \mathbf{P}_j \tag{3.3}$$

where the m_{ij} form a 3×3 matrix of constants. Each of the $\mathbf{Q}_i$ can be represented in terms of the $\mathbf{P}_j$. Substituting Equation (3.3) into Equation (3.2), we obtain

$$\mathbf{C} \sim \bigoplus_{i=1}^{3} q_i \bigoplus_{j=1}^{3} m_{ij} \mathbf{P}_j \tag{3.4}$$

$$\mathbf{C} \sim \bigoplus_{j=1}^{3} p_j \mathbf{P}_j \tag{3.5}$$

where the p_j are constants. Thus,

$$p_j = \sum_{i=1}^{3} m_{ij} q_i. \tag{3.6}$$

Therefore,

$$\begin{bmatrix} p_1 \\ p_2 \\ p_3 \end{bmatrix} = \begin{bmatrix} m_{11} & m_{21} & m_{31} \\ m_{12} & m_{22} & m_{32} \\ m_{13} & m_{23} & m_{33} \end{bmatrix} \begin{bmatrix} q_1 \\ q_2 \\ q_3 \end{bmatrix}. \tag{3.7}$$

The matrix m_{ij} can be inverted to convert from p_j to q_i.

It is convenient to specify colours using the tristimulus values of the colour stimulus entering the eye. There is currently a fair understanding of the response functions of the light-sensitive pigments in the retina of the eye; it is possible to correct for some of the simpler variables affecting this internal stimulus, but for measurement purposes it is more convenient not to perform these corrections.

Tristimulus values are defined using primaries with colour-matching functions. We will derive the form of these functions. For a given stimulus $\mathbf{E}$, the corresponding spectral energy distribution $E(\lambda)$ can be approximated with an additive mixture of n distinct monochromatic sources $\mathbf{Q}_i$ with frequency λ_i such that $\lambda_{i-1} < \lambda_i$ for all i. Let $\triangle\lambda_i = \lambda_i - \lambda_{i-1}$ define interval i. The spectral energy of $\mathbf{E}$ at the frequency of a given monochromatic source λ_i is

$$\mathbf{E}_i = \mathbf{Q}_i E(\lambda_i) \tag{3.8}$$

where $\mathbf{Q}_i$ are the tristimulus values of the source. The stimulus $\mathbf{E}$ can be approximated by

$$\mathbf{E} \approx \sum_{i=1}^{n} \mathbf{E}_i \triangle\lambda_i = \sum_{i=1}^{n} \mathbf{Q}_i E(\lambda_i) \triangle\lambda_i. \tag{3.9}$$

Now, each monochromatic source $\mathbf{Q}_i$ can be expressed as a mixture of a given set of three primaries $\mathbf{X}_1$, $\mathbf{X}_2$, $\mathbf{X}_3$:

$$\mathbf{Q}_i = x_{i1}\mathbf{X}_1 + x_{i2}\mathbf{X}_2 + x_{i3}\mathbf{X}_3. \tag{3.10}$$

Substituting Equation 3.10 into Equation 3.9, we obtain

$$\mathbf{E} \approx \sum_{i=1}^{n} \left[x_{i1}\mathbf{X}_1 + x_{i2}\mathbf{X}_2 + x_{i3}\mathbf{X}_3\right] E(\lambda_i) \triangle\lambda_i. \tag{3.11}$$

Taking the limit of this equation, we arrive at

$$\mathbf{E} = \lim_{\triangle\lambda_i \to 0} \left[x_{i1}\mathbf{X}_1 + x_{i2}\mathbf{X}_2 + x_{i3}\mathbf{X}_3\right] E(\lambda_i)\triangle\lambda_i \tag{3.12}$$

$$\mathbf{E} = \int \left[x_{i1}\mathbf{X}_1 + x_{i2}\mathbf{X}_2 + x_{i3}\mathbf{X}_3\right] E(\lambda_i)d\lambda \tag{3.13}$$

$$\begin{aligned} \mathbf{E} &= \int x_{i1}\mathbf{X}_1 E(\lambda_i)d\lambda + \int x_{i2}\mathbf{X}_2 E(\lambda_i)d\lambda & (3.14)\\ &\quad + \int x_{i3}\mathbf{X}_3 E(\lambda_i)d\lambda & (3.15) \end{aligned}$$

We can also express $\mathbf{E}$ in term of the primaries: $\mathbf{X}_1$, $\mathbf{X}_2$, $\mathbf{X}_3$:

$$\mathbf{E} = x_1\mathbf{X}_1 + x_2\mathbf{X}_2 + x_3\mathbf{X}_3. \tag{3.16}$$

Matching terms in the previous two equations gives the result as the form for colour-matching equations:

$$x_i = \int E(\lambda)\bar{x}_i(\lambda)d\lambda. \tag{3.17}$$

Two tristimulus coordinate systems in wide use are the *RGB* and *XYZ* primary systems[WS82]. The colour matching functions for each of the three primaries must be determined empirically for a population of observers. The conversion of tristimulus values with a matrix multiplication implies the conversion of the corresponding colour matching functions with a matrix multiplication is also possible.

In the CIE 1931 *RGB* Primary System, or *RGB*-system, the primary colours defined by the three chosen colour matching functions are realizable by monochromatic colours. Specifically, the colour matching function R (red) is chosen to correspond to the sensitivity of a standard observer to 700 nm light; G (green), 546.1 nm; and B (blue), 435.8 nm. The *tristimulus values* of colours in this system are specified in the order (R, G, B) and specify the required mixture of the three primaries

for a colour match. *Chromaticity coordinates* are specified in the order (r, g, b) and are obtained by normalizing the tristimulus values as follows:

$$r = \frac{R}{R+G+B} \tag{3.18}$$

$$g = \frac{G}{R+G+B} \tag{3.19}$$

$$b = \frac{B}{R+G+B} \tag{3.20}$$

The CIE 1931 XYZ Primary System, or XYZ-system, is defined by a set of three colour-matching functions $\bar{x}(\lambda), \bar{y}(\lambda), \bar{z}(\lambda)$, which also define the CIE 1931 standard colorimetric observer. These colour-matching functions are specified in terms of spectral intensity over the visible region of the spectrum. The chromaticity coordinates of this system are specified in the order (x, y, z) and are obtained by normalizing the tristimulus values as follows:

$$x = \frac{X}{X+Y+Z} \tag{3.21}$$

$$y = \frac{Y}{X+Y+Z} \tag{3.22}$$

$$z = \frac{Z}{X+Y+Z} \tag{3.23}$$

The relation between r, g, b and x, y, z chromaticity coordinates is:

$$x = \frac{0.49000r + 0.31000g + 0.20000b}{0.66697r + 1.13240g + 1.20063b} \tag{3.24}$$

$$y = \frac{0.17697r + 0.81240g + 0.01063b}{0.66697r + 1.13240g + 1.20063b} \tag{3.25}$$

$$z = \frac{0.00000r + 0.01000g + 0.99000b}{0.66697r + 1.13240g + 1.20063b} \tag{3.26}$$

The XYZ primaries are related to the RGB primaries by the following set of equations.

$$R = 0.73467X + 0.26533Y + 0.0Z \tag{3.27}$$

$$G = 0.27376X + 0.71741Y + 0.00883Z \quad (3.28)$$

$$B = 0.16658X + 0.00886Y + 0.82456Z \quad (3.29)$$

Note that R is not a functions of Z. The conversion between the colour-matching functions of the two systems is not given here as it is not directly relevant to the implementation of OSA PlaneSight. The three primary colours implicit in the definition of the *XYZ*-system are not realizable by actual colours. In the *XYZ*-system, the vectors that define the primary colours were chosen such that the chromaticity values for all realizable colours would have positive values for (x, y, z).

Chromaticity diagrams have the useful property that additive mixtures of any two lights fall on straight lines on the diagram; this follows from the projective nature of the definition of the chromaticity coordinates. There are many perceivable colours that have a negative r value when projected on an (r, g)-chromaticity diagram; this is inconvenient when computing tristimulus values from a given spectral radiant power distribution and the standard colour-matching functions. Another fault with the (r, g) diagram is that there exist values for which $R + G + B = 0$. These problems are not the case on a (x, y)-chromaticity diagram.

For large-field colour matching, it was found that the CIE 1931 standard observer was inadequate. The primary motivation for a new standard better-suited for large-field colour matching was that as the size of the field of view is changed, there is a continuous change in the amounts of the primaries required for a colour match due to the structure of the retina[PS76]. The CIE 1931 standard observer was derived for a 2°-field observer, and thus not well-suited to large-field colour matches. A 10°-field observer was proposed to facilitate the exploration of such colour matches.

The three colour matching functions $\bar{x}_{10}(\lambda), \bar{y}_{10}(\lambda), \bar{z}_{10}(\lambda)$ are used to define the CIE 1964 standard observer. This (X, Y, Z)-system is intended for practical

colorimetry involving colour matching-field which subtend angles of greater than 4°.

The *RGB* colour-matching functions for the 1964 observer were chosen to correspond to 645.2 nm for R (red), 526.3 nm for G (Green), and 444.4 nm for B (Blue). The conversion between the two sets of colour-matching functions is:

$$\bar{x}_{10}(\lambda) = 0.341080\bar{r}_{10}(\lambda) + 0.189145\bar{g}_{10}(\lambda) + 0.387529\bar{b}_{10}(\lambda), \quad (3.30)$$

$$\bar{y}_{10}(\lambda) = 0.139058\bar{r}_{10}(\lambda) + 0.837460\bar{g}_{10}(\lambda) + 0.073316\bar{b}_{10}(\lambda), \quad (3.31)$$

$$\bar{z}_{10}(\lambda) = 0.000000\bar{r}_{10}(\lambda) + 0.039553\bar{g}_{10}(\lambda) + 2.026200\bar{b}_{10}(\lambda). \quad (3.32)$$

Notation for tristimulus values and chromaticity coordinates in the CIE 1964 observer are analogous to those of the CIE 1931 observer; the symbols for each tristimulus value and chromaticity coordinate are subscripted with the number 10. There is no linear relation between the chromaticity coordinates or colour matching functions of the CIE 1964 and CIE 1931 observers.

3.5 Uniform Colour Spaces

Ideally, a perceptually uniform (three-dimensional) colour space has the property that the magnitude of perceptual differences between two colours corresponds to the distance between the two colours in that colour space. Such a uniform colour space and accompanying colour difference formula simplifies the problem of specifying colours, colour tolerances, and colour differences. As well, it can provide a basis for the selection of harmonious colour combinations. The result of strategies used in the selection of a harmonious set of colours often forms geometric patterns when placed in a lattice of a uniform colour space.

Many colour spaces can be characterized by a *metric*, a function defining the geometry of the colour space. It is advantageous to have a colour space where the perceptual difference between two colours can be calculated from the distance between the two colours. Such formulae are called *difference formulae*. Ideally, a uniform colour space is Euclidean. Cartesian coordinates are preferred, as colour difference in human vision can be approximated using a Cartesian distance formula. The difference formula for Cartesian space is simple to determine and evaluate, though some attempts at uniform spaces have used other coordinate systems. To date, only approximately uniform colour spaces have been proposed. There is evidence that a true perceptually uniform colour space does not exist [WF71].

Two approximately uniform colour spaces and corresponding colour difference formulae recommended by the CIE are the CIE 1976 ($L^*u^*v^*$)-space or *CIELUV*, and the CIE 1976 ($L^*a^*b^*$)-space or *CIELAB*[WS82]. Both systems are currently in widespread use. The choice of which system is used depends on the requirements of the application, as both systems offer approximately uniform colour spacing. L^* is a measure of for lightness, or value. The formula for L^*, which is the same in both systems, was chosen to be a cube-root expression rather than the more complex quintic expression for value used in the Munsell system[Rob90], which is discussed briefly in section 3.7.

The *CIELUV* coordinates L^*,u^*,v^* are defined as:

$$L^* = 116\left(\frac{Y}{Y_n}\right)^{\frac{1}{3}} - 16 \tag{3.33}$$

$$u^* = 13L^*(u' - u'_n) \tag{3.34}$$

$$v^* = 13L^*(v' - v'_n) \tag{3.35}$$

provided that $Y/Y_n > 0.01$. For $Y/Y_n \leq 0.008856$ a different formula is used:

$$L_m^* = 903.3\frac{Y}{Y_n}. \tag{3.36}$$

Both *CIELUV* and *CIELAB* are undefined in the interval $(0.008856, 0.01)$. The variables u', v', u'_n, v'_n are determined by

$$u' = \frac{4X}{X + 15Y + 3Z} \tag{3.37}$$

$$v' = \frac{9Y}{X + 15Y + 3Z} \tag{3.38}$$

$$u'_n = \frac{4X_n}{X_n + 15Y_n + 3Z_n} \tag{3.39}$$

$$v'_n = \frac{9Y_n}{X_n + 15Y_n + 3Z_n} \tag{3.40}$$

where X_n, Y_n, Z_n are the tristimulus values of the "white" object-colour stimulus as determined by the illuminant. The *CIELUV* space has a chromaticity diagram (u', v'). A u^*, v^*-diagram is not a chromaticity diagram, since straight lines in XYZ are generally curved when projected on such a diagram.

The colour difference formula for *CIELUV* is:

$$\Delta E_{uv}^* = \left[(\Delta L^*)^2 + (\Delta u^*)^2 + (\Delta v^*)^2\right]^{\frac{1}{2}}. \tag{3.41}$$

In industry, it is sometimes desirable to weight the components differently.

The *CIELAB* coordinates L^*, a^*, b^* are defined as:

$$L^* = 116\left(\frac{Y}{Y_n}\right)^{\frac{1}{3}} - 16 \tag{3.42}$$

$$a^* = 500\left[\left(\frac{X}{X_n}\right)^{\frac{1}{3}} - \left(\frac{Y}{Y_n}\right)^{\frac{1}{3}}\right] \tag{3.43}$$

$$b^* = 200\left[\left(\frac{Y}{Y_n}\right)^{\frac{1}{3}} - \left(\frac{Z}{Z_n}\right)^{\frac{1}{3}}\right] \tag{3.44}$$

for X/X_n, Y/Y_n, $Z/Z_n > 0.01$. Somewhat modified formulae for a^* and b^* are required for values of X/X_n, Y/Y_n, Z/Z_n that are less than or equal to 0.008856. The following formulae are then recommended:

$$L^*_m = 903.3\frac{Y}{Y_n}, \quad \frac{Y}{Y_n} \leq 0.00856 \tag{3.45}$$

$$a^*_m = 500\left[f\left(\frac{X}{X_n}\right) - f\left(\frac{Y}{Y_n}\right)\right] \tag{3.46}$$

$$b^*_m = 200\left[f\left(\frac{Y}{Y_n}\right) - f\left(\frac{Z}{Z_n}\right)\right] \tag{3.47}$$

where

$$f(x) = \begin{cases} 7.787x + \frac{16}{116}, & x \leq 0.008856 \\ x^{\frac{1}{3}}, & x > 0.008856 \end{cases} . \tag{3.48}$$

The *CIELAB* space is attractive because the colour difference formula is similar to that of the Adams-Nickerson formula which had already been used widely in industry when the proposed standard was being deliberated. The *CIELAB* system is often used when small colour differences are being studied. Some systems use the *CIELAB* formulae as a base, adding correction formulae and terms to better fit visual data.

The colour difference formula for *CIELAB* is:

$$\Delta E^*_{ab} = \left[(\Delta L^*)^2 + (\Delta a^*)^2 + (\Delta b^*)^2\right]^{\frac{1}{2}} . \tag{3.49}$$

CIELUV is undefined for values of Y between 0.008856 and 0.01. *CIELAB* is undefined for values of X, Y, and Z between 0.008856 and 0.01. We were unable to determine reason for this from the literature referenced.

Another uniform colour space, proposed by the Optical Society of America, is discussed separately in section 3.8.

3.6 Colour Constancy and Surface Perception

In the real world, the viewing conditions in which humans observe objects vary dramatically when compared with carefully controlled laboratory conditions. Many factors are involved, such as the sources of illumination, the geometric arrangement of illumination with respect to the objects being illuminated, the surface texture of the illuminated objects, glare, and the position of the observer. For a colour to be useful in identifying an object, whether on a workstation display or on a dinner table, the colour must appear to remain roughly constant under different viewing conditions. Essentially, the phenomenon of *colour constancy* involves the colour appearance of objects being maintained despite variations in the colour of nearby objects, the spectral distribution of ambient illumination, the angle of view, and other changes in viewing conditions. Though colour constancy is not perfectly maintained in human colour vision, it is approximated under a wide range of viewing conditions.

There has been much research on the phenomenon of colour constancy and models of colour constancy. Perceived colour depends on the spectral power distribution of the ambient light and the surface reflectance of the objects being observed, as well as the geometry of the scene. The appearance of an object in a scene is somewhat better predicted by the reflectances of the surfaces involved than by light actually entering the eye. The colour of an object must be recognizable under varying viewing conditions in order to be useful in carrying information. Therefore the human visual system does not look for specific colours, but seeks out colours corresponding to desired object reflectances. Ideally, a colour constancy algorithm takes input in the form of receptor responses and computes colour values dependent only on the surface reflection spectra. The human visual system is believed

to operate in a manner similar to this. As a demonstration of colour constancy, Land[Lan59] was able to generate an imperfect, but fair rendition of colour using only white and red light. Land photographed a scene on black-and-white film, once using a red filter and once using a green filter. He projected white light through the red-filtered image through a red filter and then superposed the image with white light shone through the green-filtered image. A naive theory of colour perception would suggest the image should seem to appear entirely as shades of pink, which was not the case.

One of the most successful algorithms for modeling colour constancy is the retinex algorithm[BW86]. The retinex theory[Lan77] hypothesizes that the retinex, the "retina-and-cortex system", treats colour in terms of codes that are based upon three-part responses from the retina. The retina generates responses by dividing the visible spectrum into overlapping high, middle, and low ranges of sensitivity. The codes generated from these responses are highly independent of the flux of radiant energy entering the eye, but are highly correlated with the object reflectances.

It is interesting to compare the retinex algorithm with a simple colour correction procedure, namely normalization to a reference reflective surface; this method is often used in industry in colour correcting video signals. Such a procedure results in a proper correction for changes in illuminant for a restricted but significant class of surface reflectance functions and spectral power distributions of ambient light.

However, studies have indicated that the retinex algorithm is inadequate as a model of human colour vision. Many people can, either under instruction or voluntarily, change their interpretation of a scene to alter their perceptions of colour constancy[Hen35, Jud40].

In general, it is not possible to separate the spectral distribution of ambient light and from surface reflectances under all viewing conditions. However, surface

reflectance is of lesser consequence for workstation CRTs under normal lighting conditions because the light from the CRT is more a function of emitted light than of ambient light. Problems are introduced when using real world viewing conditions; these are briefly mentioned in Chapter 4.

Colour constancy has been hard to model due to the difficulty of quantifying colour constancy relationships[Bri88]. Colour constancy can be violated by *metamerism* and *paramerism*. Two reflective surfaces are said to be *metameric* when they produce identical visual responses under one light source but different responses under another. Two reflective surfaces are termed *parameric* when they appear indistinguishable under one light source while having different appearances under another; the colours are not necessarily identical in response, but the colour difference is too small to be perceived by the human vision system.

3.7 Reflective Colour Systems

Material colour standards are often desirable, particularly in industry[Eva48]. These offer the convenience of being able to compare the colour of an object directly against a sample under certain lighting conditions, rather than requiring a full radiometric measurement. This approach is not immune to the effects of metamerism and paramerism, though extreme effects can be avoided or ameliorated by a proper choice of material standard for the desired application.

One set of colour standards that has had a long history of use is the Munsell System, the most recent version of which is the Munsell Renotation System. The three attributes assigned to the discrete colour samples are *hue*, *chroma*, and *value*, roughly corresponding to hue, saturation, and lightness. The published atlas, the Munsell Book of Color, contains over two thousand distinct samples, arranged in

charts containing colours of the same hue, varying in chroma and value. Charts of constant value or chroma are possible, but are infrequently used. The borders of the colour solid described by the Munsell System are irregular, as they are limited by the availability of pigments. The system is intended for use in daylight. The spacing between the colours is based on data from experiments studying human colour perception.

Another colour standard of interest is the Ostwald System. Unlike the irregular panels of the Munsell System, the borders of the panels of constant hue form triangles. The colour samples are placed in a regular triangular pattern on each panel. The panels all share the grey axis on their longest side. The resulting colour solid is enclosed by the form of two cones of equal size sharing the same base. The layout of the Ostwald System is based on the results of additively mixing colour stimuli. Colours are positioned on colour wheels of constant lightness, with the grey axis passing through the center, such that complementary colours are positioned opposite to each other about the grey axis. Series of samples parallel to the grey axis differ only in lightness. These relations are advantageous when exploring colour relations. Note that the restriction of maximum purity to a circle results in variations in saturation on the edge of wheels of constant lightness[Com63].

These colour systems offer reflective surface colour systems that are approximately uniform. These standards address larger scale colour differences than do the CIE uniform colour spaces. Material standards are not perfect, as the surface attributes of the material standard can make colour comparison difficult when the surface in question differs significantly.

3.8 The OSA Uniform Colour System

The *OSA Colour System* is not only uniformly-spaced but also highly regular in arrangement. The arrangement and its advantages are discussed below. The colorimetric basis of the OSA Colour System is presented.

The colour samples which make up the OSA Colour System are placed in a regular rhombohedral lattice in three-dimensional space. This lattice is the logical generalization of the regular triangular lattice in the plane to three-dimensional space. The lattice is superposed on a three-dimensional Cartesian grid for simplicity in specifying the coordinates of a sample point. We refer below to the OSA Colour System as the OSA space.

The three coordinates L, j, and g stand for lightness, yellowness, and greenness respectively. The notation j was chosen from the French word for yellow, *jaune*, since y for yellowness would have been confused with the CIE chromaticity coordinate y in the CIE XYZ system,

The geometric arrangement of the OSA space is illustrated in Figure 3.1. Because of the regular rhombohedral arrangement, each point enclosed within the gamut of a set of sample points is surrounded by twelve perceptually equidistant colours. The set of 424 colour sample points, known as the full-step set, forms the gamut of the OSA space. Each sample in the full-step set is specified with integer values for L, j, g such that if L is odd, then both j and g are odd and if L is even, then both j and g are even. The distance between points in the OSA space is determined by the metric $(\sqrt{2}L, j, g)$, allowing integer specification of the coordinates. If the sample point at the center of Figure 3.1 is at $(0, 0, 0)$, its twelve nearest neighbours in the full-step set are at $(0, 0, 2)$, $(0, 0, -2)$, $(0, 2, 0)$, $(0, -2, 0)$, $(1, 1, 1)$, $(1, 1, -1)$, $(1, -1, 1)$, $(1, -1, -1)$, $(-1, 1, 1)$, $(-1, 1, -1)$, $(-1, -1, 1)$, and

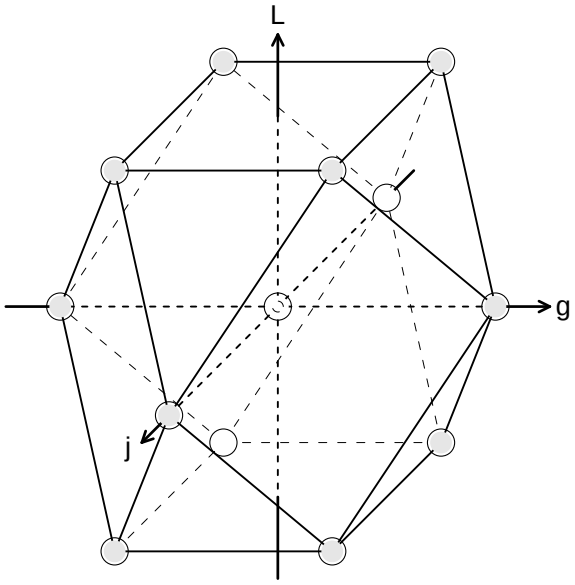

Figure 3.1: Schematic diagram of cubo-octahedron, illustrating packed-sphere structure of the OSA lattice.

$(-1, -1, -1)$.

In Figure 3.1, the sample points are represented by small circles. The twelve nearest neighbours of the central point describe a polyhedron called a *cubo-octahedron*. If this polyhedron were solid, the shaded circles would be visible, while the clear circles would be occluded, or in the case of the center circle, enclosed. Dotted lines are added to clarify the shape of the cubo-octahedron; if the polyhedron were solid, the solid lines connecting two sample points would be visible, while the thin dotted lines would be occluded. The three axes correspond to the three coordinates of the OSA space; the portion of these axes that are enclosed in the cubo-octahedron are rendered as thick dotted lines, while the portion outside is drawn as thick solid lines. A physical model of a cubo-octahedron can be constructed with spheres.

We have seen above that each sample point in the OSA space is arranged such that a maximal number of points have a large number of equidistant neighbours. The regularity of this arrangement also allows for several planes passing though a given point to contain several other sample points. These symmetries can be employed in a navigation paradigm to maneuver about the OSA space. Seven planes defined in the OSA space pass through each sample point. Three of the planes contain square lattices, while the four remaining planes contain triangular lattices. Their relative orientations can be observed in the cubo-octahedron, which has six square faces and eight triangular faces. The square planes are specified by the constant equations $L = const$, $g + j = const$, and $g - j = const$; while the triangular planes are specified by $L + g = const$, $L - g = const$, $L + j = const$, and $L - j = const$. An illustration of a triangular lattice plane is given in Figure 3.2. The planes can be described with simple equations. In Figure 3.3, the normals are given for three planes intersecting different facets of the cubo-octahedron.

Constant values of lightness (L) define planes of constant lightness. The coordinates of the colour space are centered on a 30% reflectance grey. Shades of grey have $j = 0$ and $g = 0$. As a result, planes which have an odd value for L do not contain a grey. Colours in the full-step set are defined for values of L between -7 and 5 inclusive. A half-step, or half-integer set of 134 colours was also defined to permit construction of colour scales with a smaller spacing than the full-step set. The half-step set encloses significantly less volume in the OSA space than the full-step set. It was decided that each colour in the OSA gamut had to be reproducible in a stable, opaque paint formulation, so that a physical set of samples of the OSA colours could be produced. The boundary of the OSA gamut is, as a result, irregular and asymmetric.

The XYZ specifications for colour samples in both the full- and half-step sets

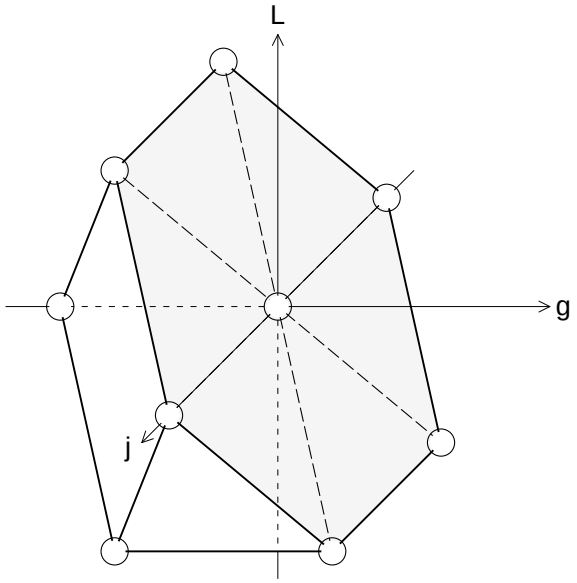

Figure 3.2: Cutaway of cubo-octahedron, illustration the orientation of triangular lattice plane $L + g = 0$.

must be derived using numerical approximation techniques. The results were given in a paper by MacAdam[Mac78]; a small number of the target XYZ specifications do not agree with the results of the OSA formulae given in [Mac74], reproduced later below.

In deriving the OSA space, colour samples were placed against a 30% reflectance grey background. The colour differences tested were all greater than twenty times *just noticeable*, the difference required to be on the perceptual threshold of colour judgement. A difference is defined as being on the perceptual threshold when it is detectable approximately 50% of the time by a standard (human) observer. Hence the OSA space should not be used in either evaluating small colour differences, or for very large colour differences. Nor are the formulae necessarily applicable outside of

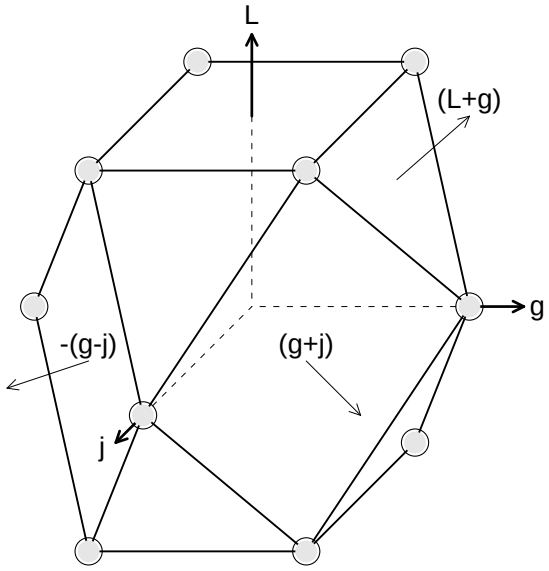

Figure 3.3: Illustration of plane orientations described by the facets of the cubo-octahedron.

the defined gamut. A set of approximately 500 painted sample cards were produced to aid in the visualization and use of the scales that can be generated using the OSA specifications. These samples provide a valuable tool for the exploration of both complementary and common-denominator colour schemes.

The highly regular and symmetric geometric arrangement of the OSA space encourages one to explore colour schemes in a manner either unlikely or impossible with other colour systems. The OSA space thus lends itself to the study of colour harmony theory in art. One artist remarked, "With the other colour systems...it is easy to locate complementary relationships among a wide range of hues. Using the OSA scales it is easy to locate far fewer complementary pairs; however...it is possible to explore complementary relationships far more thoroughly than is

possible with any other system."[Hed88] One can define such a grid and impose it on another (approximately) uniform colour space, such as *CIELUV*. However, the choice of an appropriate spacing and therefore colour difference between points in the grid is not readily apparent. The choice of spacing for the OSA Colour System is suggested by the data used in its derivation.

The coordinates of the OSA Colour System, L,j,g, were originally defined as[1]:

$$L = \frac{5.9\left[\bar{Y}_{10}^{\frac{1}{3}} - \frac{2}{3} + 0.042(\bar{Y}_{10} - 30)^{\frac{1}{3}}\right] - 14.4}{\sqrt{2}} \tag{3.50}$$

$$j = C(1.7R_{10}^{\frac{1}{3}} + 8G_{10}^{\frac{1}{3}} - 9.7B_{10}^{\frac{1}{3}}) \tag{3.51}$$

$$g = C(-13.7R_{10}^{\frac{1}{3}} + 17.7G_{10}^{\frac{1}{3}} - 4B_{10}^{\frac{1}{3}}) \tag{3.52}$$

where

$$R_{10} = 0.7990X_{10} + 0.4194Y_{10} - 0.1648Z_{10} \tag{3.53}$$

$$G_{10} = -0.4493X_{10} + 1.3265Y_{10} + 0.0927Z_{10} \tag{3.54}$$

$$B_{10} = -0.1149X_{10} + 0.3394Y_{10} + 0.7170Z_{10} \tag{3.55}$$

$$\begin{aligned}\bar{Y}_{10} = {} & Y_{10}(4.4934x_{10}^2 + 4.3034y_{10}^2 - 4.276x_{10}y_{10} \\ & -1.3744x_{10} - 2.5643y_{10} + 1.1803)\end{aligned} \tag{3.56}$$

and

$$C = 1 + 0.042\frac{(\bar{Y}_{10} - 30)^{\frac{1}{3}}}{\bar{Y}_{10}^{\frac{1}{3}} - \frac{2}{3}} \tag{3.57}$$

[1]There was some difficulty in determining the correct OSA equations, as the Wyszecki and Stiles book on colour science[WS82] incorrectly reproduced the formulae derived from the original paper by MacAdam[Mac74]; the formulae for j and g were reversed and the normalization factor $L = (\mathrm{L} - 14.4)/\sqrt{2}$ omitted. This normalization factor is easily overlooked in the MacAdam paper because of its placement, and has been inadvertently omitted in at least one paper[MG80].

hold. For a perfect reflecting diffuser illuminated by CIE standard illuminant D_{65}, the "white" object-colour stimulus is $\bar{Y}_{10} = 100$. The corresponding difference formula is:

$$\left[2(\Delta L^*)^2 + (\Delta j^*)^2 + (\Delta g^*)^2\right]^{\frac{1}{2}}. \tag{3.58}$$

The OSA equations are stated above as a conversion from a CIE XYZ specification to corresponding coordinates in the OSA colour space: L,j,g. The inverse transformation has no solution in terms of elementary functions. A numerical inverse can be computed using a multivariable form of Newton's method, which we will now describe.

Newton's method is a general algorithm to determine roots for a system of equations $\mathbf{F}(\vec{x}) = 0$. An initial approximation $\vec{x}$ is required as input. The approximation step is $\vec{x}' = \vec{x} + \vec{y}$, where $\vec{y}$ is determined by the equation $J(\vec{x})\vec{y} = -\mathbf{F}(\vec{x})$. The Jacobian matrix $J(\vec{x})$ is a matrix whose elements are defined as $J(\vec{x})_{i,j} = \frac{\delta f_i(\vec{x})}{\delta x_j}$. It is straightforward to derive and evaluate the partial derivatives that form the Jacobian matrix; the resulting linear system of equations is easily solved by methods such as Gaussian elimination to obtain $\vec{y}$. The result $\vec{x}'$ can be used as the initial approximation for a further approximation step. Iterations of the approximation step can be repeated until the absolute value of $\vec{y}$ falls within the desired precision, provided that the system of equations converged about the desired region.

When calculating the inverse for the OSA Colour System transformation, the appropriate system of equations is $\mathbf{F}(\vec{x}) = \mathbf{G}(\vec{x}) - \vec{k}$, where $\mathbf{G}(\vec{x})$ transforming an XYZ coordinate to L,j,g, for a desired OSA coordinate $\vec{k}$ and an XYZ approximation $\vec{x}_0$. In this case, $\vec{x} = (X, Y, Z)$, $F_1(\vec{x}) = L(\vec{x})$, $F_2 = j(\vec{x})$, and $F_3 = g(\vec{x})$; the

Jacobian matrix is:

$$J(\vec{x}) = \begin{bmatrix} \frac{\delta L}{\delta X} & \frac{\delta L}{\delta Y} & \frac{\delta L}{\delta Z} \\ \frac{\delta j}{\delta X} & \frac{\delta j}{\delta Y} & \frac{\delta j}{\delta Z} \\ \frac{\delta g}{\delta X} & \frac{\delta g}{\delta Y} & \frac{\delta g}{\delta Z} \end{bmatrix}. \tag{3.59}$$

The j and g coordinates tend to converge much faster than the L coordinate by this method due to the curvature of the OSA colour space.

The choice of the initial approximation point is of some concern, as the OSA colour space is not well-behaved. A poor choice of initial approximation will increase the number of iterations required for a solution. In certain cases, the algorithm will not converge. Application of the OSA space equations in an extrapolation outside the defined OSA gamut revealed a significant amount of curvature and numerical instability. It is apparent that the equation for L alone does not hold when the colour in question is far from the grey line, i.e. when $\sqrt{j^2 + g^2} > 20$, a region that is well outside the bounds of the experiments that provided the data used in deriving the OSA space. This can be an obstacle to extending use of the equations to the full CRT monitor gamut if the monitor model is not chosen carefully.

Recently, several parameters of the OSA Colour System formulae were redetermined by regression after a further study employing the OSA set of colour cards, obtaining judgement data for "colour differences that are distributed much more uniformly in colour space than in the committee's [original] study and to include a much larger proportion of combinations of both lightness and chromatic contrast."[Mac90] The purpose of this redefinition was to provide a better fit between the OSA formulae and the data from perceptual experiments. No revised OSA full- or half-step sets of sample points were given however.

The chromaticity of the green primary was changed. The resulting primary chromaticities are $x_R = 0.747$, $x_R = 0.253$; $x_G = 0.945$, $y_G = -1.522$; $x_B = 0.171$,

$y_B = 0$. The formulae for R, G, B and L, j, g were revised for the OSA space as follows:

$$R = 0.9285X_{10} + 0.3251Y_{10} - 0.1915Z_{10} \quad (3.60)$$

$$G = -0.4493X_{10} + 1.3265Y_{10} + 0.0927Z_{10} \quad (3.61)$$

$$B = -0.2032X_{10} + 0.6Y_{10} + 0.5523Z_{10} \quad (3.62)$$

$$L = \frac{5.9\left[\bar{Y}_{10}^{\frac{1}{3}} - \frac{2}{3} + 0.06(\bar{Y}_{10} - 30)^{\frac{1}{3}}\right] - 14.4}{\sqrt{2}} \quad (3.63)$$

$$j = C(-1.3R^{\frac{1}{3}} + 17G^{\frac{1}{3}} - 15.7B^{\frac{1}{3}}) \quad (3.64)$$

$$g = C(-12.7R^{\frac{1}{3}} + 19G^{\frac{1}{3}} - 6.3B^{\frac{1}{3}}). \quad (3.65)$$

Chapter 4

Calibrated Colour on the Workstation CRT

4.1 Introduction

This chapter begins with a discussion of the hardware of the shadowmask CRT and the mechanism by which it produces light. A model of this mechanism will be used in calibration.

A suitably accurate means of calibrating the CRT is required for accurate portrayal of the OSA-UCS space colours. Such a calibration is required to ensure that the colours rendered correspond to and exhibit the properties associated with the OSA space colours. Several approaches to calibrating a CRT are considered, including their advantages and disadvantages, and a choice of calibration model for OSA PlaneSight is justified.

4.2 The Shadowmask CRT

The shadowmask cathode ray tube (CRT) is currently the dominant colour video display technology[Cow89b]. The tube of a CRT is a sealed glass tube from which all air has been evacuated. Inside this vacuum tube are three *electron guns*, which consist of the following pieces. First, cathodes which are heated until they emit electrons. Surrounding each cathode is a *control grid* containing a hole through which some of the emitted electrons pass; the number of electrons passing through the hole is determined by the voltage applied to the control grid, which is exactly the voltage level specified by the relevant digital-to-analog converter (DAC). These emitted electrons are then focussed into beams and accelerated toward the faceplate where they strike at positions determined by the appropriate deflection circuitry. Light is produced when they excite phosphors on the inner surface of the CRT screen; near and inside this *faceplate* is a metal screen called a *shadowmask*. The shadowmask is perforated with a regular pattern which allows it to either stop the beams by conducting them away or allowing them to pass through to strike the phosphor dots on the back of the faceplate. Several different geometries are in common use, but all allow each of the three guns to irradiate only one of the three coloured phosphors; these are generally said to be red, green, and blue, though their exact chromaticities vary from one monitor model to another and each emits light over a rather broad portion of the spectrum.

Next we will discuss several characteristics of the CRT that affect calibration, including deviations from gamma correction, misconvergence, lack of purity, and stray magnetic fields.

The nonlinear relation between the input voltage applied to a CRT and the intensity of the light emitted from the screen, measured experimentally, is generally

of the form

$$\phi = \phi_0(V/V_0)^\gamma \tag{4.1}$$

where ϕ is the current light intensity, ϕ_0 is the maximum light intensity, V is the current input voltage, V_0 is the maximum input voltage, and γ is the correction exponent. The constant γ suggests the term *gamma correction*, used to describe the relation. The actual relation is closely modelled over a large range, but for extreme input voltages there are significant deviations.

Convergence describes the process by which the three guns strike the same point on the tube at the same time. When the guns do not converge properly, there are spurious fringes of colour at edges where intensity changes significantly. Convergence is usually set permanently at the factory. When convergence is poor, it is often due to small, stray magnetic fields, such as those produced by neighboring electronic equipment or even the earth.

Purity is a measure of the degree to which inappropriate phosphors are caused to emit by each electron beam. The purity of a monitor does not need to be adjusted in general, though it is affected by stray magnetic friends.

Magnetic fields build up in the shadowmask over time. To combat these unwanted fields, a set of wires is usually built in to the monitor around the edge of the faceplate that is used in *degaussing* the shadowmask. This involves passing a current through the wires, removing the magnetic fields that have built up. This is done automatically at power-up, and can be manually performed if a degaussing control has been provided and the monitor is on.

The colour produced when all phosphors are emitting light at maximum intensity is referred to as a *calibration white*, which has an associated chromaticity. When the same input voltage is applied to the red, green, and blue guns, it is desirable

that the same chromaticity be produced at all voltage levels. The gamma factors therefore should be identical for all three phosphor guns.

4.3 Colour CRT Calibration

We now discuss CRT *calibration*, a relation that describes the monitor tristimulus values resulting from given input voltages. Several approaches to calibration are introduced. Their advantages and disadvantages are considered in a justification of the calibration method implemented in OSA PlaneSight.

In general, CRT calibration models attempt to parameterize phosphor output in terms of inputs and parameterized equations. CRT calibration requires radiometric measurement of the phosphor output for a given a set of gun voltages. Each calibration model has different parameters and requires different measurements.

The process of radiometric measurement requires the use of a *monochromator* and a *radiometer*. A *monochromator* is a device which takes radiant energy and splits it into component spectral parts, which are measured separately. Either a prism or a diffraction grating[1] may be used as a monochromator. A *radiometer* measures a part of the spectrum isolated by the monochromator. A *spectroradiometer* combines the functions of both radiometer and monochromator. The technical details of such hardware are quite simple, but beyond the scope of this paper. Current technology allows the automation of radiometric measurement.

There are several peculiarities of the CRT which must be taken into account

[1]When light passes by the edge of an opaque body or a narrow slit, parallel, alternating fringes of light and dark bands are produced. Diffraction gratings use this effect to allow measurement of wavelength. Such gratings are commonly made by scoring equally spaced, parallel grooves on a glass or plastic plate.

when performing radiometric measurement. These are discussed at length in the literature [Cow89b, Bra89]. For instance, for a fixed combination of input values, the intensities of the three phosphors vary significantly between different locations on the monitor. The light emitted by the monitor also varies over time as the screen is being refreshed.

The measurements made are of little use over extended periods if, for instance, the phosphor chromaticities change significantly. The *temporal stability* of a monitor refers to the stability of the parameters relevant to calibration. If these parameters vary too much, the assumptions made during the calibration will not hold, and the calibration will be invalid. For instance, phosphor luminances tend to decline over a period of years. Fortunately, the chromaticity drift in a CRT tends to be relatively small over time, as the intensity drifts of the different phosphors are of approximately the same magnitude[Bra89].

No single type of calibration method is suitable for all applications. There are three major types: exhaustive, local, and model-dependent. Each has its strengths and weaknesses. An *inverse calibration* relates monitor inputs from given tristimulus values; the ease of determining this relation varies from model to model. Colours that lie outside the range of produced by a given monitor, *out-of-gamut* colours, can be easy, difficult, or impossible to determine under a given model.

Exhaustive calibration methods allow high precision over the entire gamut of the monitor, but are very expensive in terms of computation and radiometric measurement. An exhaustive calibration method requires the measurement of all the colours that can be produced by the monitor and the construction of a table that translates monitor coordinates (generally R, G, B) to their corresponding tristimulus values. Truly exhaustive methods are often impractical. For example, 24-bit colour (8 bits assigned to each gun) allows 16,777,216 different monitor outputs; if

one colour is measured per second, an exhaustive calibration would take almost 200 days. In practice, a subset of the available colours is chosen for measurement and remaining colours are interpolated. Algorithms written for exhaustive calibration can be applied to other output devices, such as printers.

The treatment of out-of-gamut colours given such calibration depends largely on the needs of the application. There is no natural solution to approximating the out-of-gamut colour. Algorithms exist for projecting the out-of-gamut colour onto the surface of the monitor gamut, although they are computationally expensive.

Inverse computations (obtaining the required voltages for a given colour) can be accomplished by applying Newton's method for three variables or by interpolating trilinearly upon the data in a partial inverse table. The former is subject to failure if the data has numerical instability due to noise. Both solutions are computationally expensive.

Local calibration methods take advantage of a need to calibrate only a small fraction of the monitor gamut. These restrictions allow certain simplifications to be made. Such methods allow high precision, but only for a small part of the monitor gamut, and are moderately expensive, both in computation and radiometric measurement. Algorithms developed with these methods can often be generalized and adapted for other output devices.

How to deal with out-of-gamut colours depends on the nature of the local calibration method used and the requirements of the application. Again, some methods do not provide for a means of handling out-of-gamut colours, while others allow for relatively easy approximation.

Some methods of local calibration do not lend themselves to inverse calculation, such as calibration of a small set of discrete colours, while others allow an inverse

to be calculated easily, such as a one-dimensional colour space.

Unlike the previous classes of calibration methods, model-dependent calibrations require a model of how the device in question produces colour. Such models are parameterized, allowing the device to be calibrated by taking only as many measurements as are needed to estimate the parameters. Model-dependent calibration methods are less expensive with respect to computation and measurement, and offer moderate precision over the entire device gamut. The most common model-dependent monitor calibration model is detailed in the next section. A model-based calibration may be useful even though some of the underlying assumptions do not hold. The ease with which inverses may be computed varies from model to model.

A model-dependent calibration method was used in the work reported here for the following reasons. Since our ideal colour selection tool must span the monitor gamut, local methods are unsuited for this task. Secondly, realtime interaction is desired. Exhaustive methods are unsuitable without a large amount of precomputation. Finally, the colour selection tool must operate under a variety of viewing conditions. The relative error tolerance of the model-dependent method was thought to be important. Exhaustive methods cannot explicitly offer such tolerance. In addition, obtaining parameters for a new monitor is accomplished far more quickly with the model-dependent method used than for an exhaustive method.

4.4 Gamma Correction: a Model-dependent Calibration Method

For the purposes of OSA PlaneSight, it is desirable to visually approximate as closely as possible the conditions of a calibrated monitor without actually per-

forming a calibration. It is impractical to expect each monitor in the intended environment to be calibrated and used under calibrated conditions comprised of standard illumination, standard viewing distance and angle, and possibly even a standard room. Monitors are viewed in the ordinary working environment under widely varying viewing conditions which change over time. It is desirable that the calibration model selected be tolerant of errors due to viewing environments and/or monitor parameters that do not match the calibration conditions. The model of the CRT underlying the chosen calibration method is described first, followed by an explanation of the parameters of the model.

The standard model of a colour CRT has four parts.

1. Colour is produced by the additive mixture of three component colours; they are almost always the light emitted by the three phosphors.

2. The intensity of light in each component is determined by a single input signal which is independent of the other two. The input signals are generally referred to as R,G,B for the red, green, and blue phosphors respectively.

3. The relative spectral power distribution of the light in each component is constant over all values of input signal voltage. Thus the chromaticity of each component colour is constant.

4. The intensity of the light in each component is a linear function of the corresponding input voltage raised to a constant power.

It was decided that the *gamma correction model* would be used to characterize the phosphor output of the colour monitor. In using the standard model of the CRT, a few assumptions are required. The first part of the standard model is

trivially satisfied by shadowmask CRTs which have three phosphors. The gamma correction equation is valid over a wide range of monitor output; the fourth part of the standard model is therefore a reasonable expectation.

For *gun independence* to hold, the light that results when all three phosphor guns are turned on together must be an additive mixture of the light emitted when each of the three guns are turned on individually. This satisfies the second part of the standard model. Gun independence depends on the design of the monitor, but holds in general. Monitors which are not designed to have gun independence can often be made to have it with relatively minor hardware modification.

For *phosphor constancy* to hold, the chromaticity (or the chromaticity coordinates) of each phosphor must be independent of the voltages applied to its corresponding gun, and of the voltages applied to the other two guns. This comprises the third part of the standard model. Empirically, phosphor chromaticities do not change significantly over time.

The parameters of this gamma correction model vary from monitor to monitor. For critical applications, they are measured for each CRT, using the following method.

First, a measurement of the ambient light of the viewing environment and the *black level* of the monitor is performed under the exact viewing conditions to be used. The black level is the output of the monitor when given minimum inputs. The measured tristimulus values are X_{0i}, where i denotes the X, Y, Z values measured. Our ideal application must be able to function in a large range of illumination conditions. The phenomenon of colour constancy, discussed earlier in Chapter 3, is relied on to resolve this conflict. Humans tend to factor out the colour of the illuminant when viewing colour in scenes. The minimum lighting in OSA PlaneSight is assumed to be absolute darkness. While inadequate from the standpoint of

calibration, this is sufficient for the needs of OSA PlaneSight, which is to preserve colour appearance.

Secondly, the *gamma exponents* of the monitor must be determined. In the gamma correction model, the values of the monitor R, G, and B are linearly proportional to the corresponding gun voltages v_a, $a = R, G, B$. The *gamma functions* of the three guns relate gun input voltage and output intensity, and can be approximated by means of the following equation for $a = R,\ G,\ B$:

$$e_a(v_a) = e_{a_{\max}} \left(\frac{v_a}{v_{a_{\max}}} \right)^{\gamma_a} \tag{4.2}$$

where $v_{a_{\max}}$ is the maximum input voltage, $e_{a_{\max}}$ is the maximum relative phosphor excitation, and γ_a is the gamma correction exponent. This equation is but a restatement of Equation 4.1 and satisfies the fourth part of the model.

The monitor *phosphor chromaticities* must also be determined. In an actual calibration, phosphor chromaticities are derived from radiometric measurement. Such a measurement can be approximated with a set of typical chromaticities, but departure from these values by the actual monitor introduces errors in the colours rendered. The consequences of such errors are dealt with in the next section.

The 3×3 matrix relating maximum monitor RGB values to XYZ values can be determined by measuring the XYZ chromaticity values of each phosphor independently. This assumes the gamma correction model holds. Let us refer to the entries as x_{ai} where a denotes a particular phosphor, and i denotes one of the x, y, z chromaticity components. Given the properties of tristimulus space as described in Chapter 3, it is then a simple matter to derive the inverse matrix converting XYZ values to RGB.

The calibration equation $E_a(v_a)$ is the product of the dependent relative exci-

tation $e_a(v_a)$ and a normalization coefficient N_a.

$$E_a(v_a) = N_a e_a(v_a) \tag{4.3}$$

The coefficients N_a are determined in the equation

$$X_i = \sum_{a=R,G,B} N_a x_{ai} e_a(v_a) + X_{i0} \tag{4.4}$$

for $i = x, y, z$. The 3×3 matrix M is defined to contain phosphor chromaticities: $m_{ai} = x_{ai}$. The inverse matrix M^{-1} is computed by conventional means. If the ambient light factor can be ignored, the above equation can be restated as

$$\sum_{i=1}^{3} X_i = N_a \cdot e_a(v_a) \tag{4.5}$$

and N_a is given by

$$N_a = \frac{1}{e_a(v_a)} \sum_{i=1}^{3} X_i. \tag{4.6}$$

In practice, the values for $e_a(v_a)$ can be approximated using the gamma correction equation or taken directly from a table of measurements. The calibration equation can be simplified by factoring the maximum relative excitation $e_{a_{\max}}$ in Equation (4.1), which are constants, into the coefficients N_a.

4.5 Calibration Simulation and Miscalibration Effects

Calibrations are performed under carefully controlled conditions. When applying a calibration to a noncritical application used in uncontrolled conditions, it is impractical to expect that each monitor will have been calibrated, or that the ambient

lighting is fixed, or that observers view the CRT from a standardized distance at a fixed angle, and so forth. By choosing a model-dependent calibration carefully, a calibration that is tolerant of errors can be achieved. This expectation is reasonable, as monitors are widely-used in uncalibrated conditions without perceptually objectionable results.

Humans readily accommodate to different illumination conditions. Humans also view television programs on maladjusted monitors without complaint or apparent dissatisfaction. If the colour distortions due to maladjusted monitors are similar in nature to those caused by changes in illuminant, then a simple reflective surface model can take advantage of colour constancy, allowing relations between colours to be preserved despite errors in actual calibration. We argue that this such distortions are similar in nature by examining several types of calibration error and comparing them to the chromaticity shifts caused by changes in illumination.

The problems of metamerism and paramerism, introduced in Chapter 3, must be avoided for colour constancy to hold. Metamerism is avoided on the monitor since colour rendition is solely in the domain of additive colour. Paramerism is avoided by the relatively distant spacing of the OSA space colours being presented. Avoiding only these conditions does not guarantee colour constancy, however.

The monitor used in presenting OSA PlaneSight was not actually calibrated. Instead, the representative chromaticity values given below of a "typical" shadow-mask CRT[Cow89b] were used. A gamma correction factor of 2.3 was assumed to be "typical".

	x	y	z
red phosphor	0.652	0.335	0.013
green phosphor	0.298	0.604	0.098
blue phosphor	0.149	0.064	0.787

The calibration white was taken to be CIE standard illuminant D_{65}, since that is most commonly used[Con80]. The chromaticity of D_{65} is $x = 0.314$, $y = 0.331$.

When a monitor differs in phosphor chromaticity from those assumed by the calibration model, then there is a nonlinear transformation of actual chromaticity coordinates from the predicted coordinates for a set of given colours. Several illustrations of such shifts are given below. We argue that the effect is similar to having a different illuminant than implied in the pseudo-calibration when the chromaticity difference is not great. Colour appearance thus tends to be preserved due to colour constancy. A few examples are presented below to demonstrate the plausibility of this claim. A thorough investigation of the phenomenon is beyond the scope of this thesis.

Monitors are generally constructed such that all three guns have the same gamma factor. If the actual gamma factor for a monitor differs from the assumed gamma factor, this results in a nonlinear transformation of the colours to be displayed. The original NTSC standard has a gamma of 2.2. A standard currently in use in Western Europe, PAL, has a gamma of 2.8. In retrospect, a gamma of 2.5 is probably a better choice, as several sources claim this is more common [Con80, Ben86, MG80].

Several (x, y)-diagrams illustrating errors resulting from the use of different monitors are presented in figures below. The heel-shaped outline indicates the range of chromaticities achieved by real colours. Large triangles inside this range indicate the limit of colour that can be produced by a given set of monitor phosphors. A dotted triangle indicates the bound of the phosphors assumed in the calibration, while a solid triangle indicates the bound of the phosphors actually used. The circles marked by crosshairs indicate the calibration whites of the monitor phosphors. The assumed calibration white is indicated by crosshairs that run parallel

to the x- and y-axes, while the actual calibration white is indicate by crosshairs set at a 45°-angle from the said axes. The vectors indicate the change from the intended chromaticity to the actual displayed chromaticity. For some of the smaller changes, the arrowheads are significantly larger than the shifts, resulting in the intended chromaticity being enclosed within the arrowhead, exaggerating the change indicated in the chromaticity diagram. These (x, y)-diagrams are accompanied by diagrams containing two views of the same data in (L^*, u^*, v^*) space. In these diagrams, the assumed calibration is for a monitor with our "typical" phosphors and a gamma of 2.3. The figures illustrate the result of using this calibration assumed by OSA PlaneSight on different monitor configurations.

Following these are diagrams which illustrate the shifts resulting from changes in illuminant. These are discussed in more detail later.

The CIE chromaticities for the NTSC standard primaries[Int86] are $x_R = 0.670$, $y_R = 0.330$; $x_G = 0.210$, $y_G = 0.710$; $x_B = 0.140$, $y_B = 0.080$. Calibration white is $x = 0.310$, $y = 0.316$.

The primaries for the Conrac graphics monitor[Con80] are $x_R = 0.628$, $y_R = 0.346$; $x_G = 0.268$, $y_G = 0.588$; $x_B = 0.150$, $y_B = 0.070$. Calibration white is CIE standard illuminant D_{65}, given earlier.

The chromaticities for a set of typical P22 phosphors[Ben86] are $x_R = 0.610$, $y_R = 0.342$, $x_G = 0.298$, $y_G = 0.588$, $x_B = 0.151$, $y_B = 0.064$. Calibration white is $x = 0.313$, $y = 0.329$.

When the chromaticity shifts for a given plane in the OSA Colour Space are projected into $L^*u^*v^*$ space, the shifted colours lie generally within a plane, though the relative spacing is skewed.

Figures 4.1, 4.2, 4.3, and 4.4 present the calculated colour differences when the

OSA PlaneSight calibration, which assumes a gamma of 2.3, is used on a monitor with gamma of 2.2. The monitor primaries are those of the "typical" monitor stated above.

In Figures 4.1 and 4.2, the colours tend to shift towards the calibration white, and away from the three points of the surrounding triangle, corresponding to the three monitor phosphor chromaticities. The colours rendered tend to be slightly more luminous than intended in the calibration because of the difference in gamma factors, though this is not visible in the (x, y)-diagrams. The colours in Figures 4.3 and 4.4 behave similarly. We argue that the pattern is analogous to the shifts that occur when ambient light is added to a scene. The effects of adding a white ambient light are illustrated later. Figure 4.4 is difficult to interpret, as the plane of colours is being viewed almost edge-on in one of the two views.

Figures 4.5, 4.6, 4.7, and 4.8 present the calculated colour differences when the OSA PlaneSight calibration, which assumes a gamma of 2.3, is used on a monitor with gamma of 2.8. Again, the OSA PlaneSight calibration is used, with "typical" phosphors are assumed in the monitor.

The general pattern of the colour shifts in these figures is away from the calibration white. The shift is more pronounced the further away the intended colour is from calibration white. The colours tend to shift towards the three points corresponding to monitor primaries. This pattern of movement is the opposite of that in the previous diagrams. Because of the difference in gamma factors, the colours rendered are also generally less luminous than expected in the calibration, though this is not apparent in the (x, y)-diagrams.

Figures 4.9, 4.10, 4.11, and 4.12 present the calculated colour differences when the OSA PlaneSight calibration is applied to a monitor with NTSC phosphor primaries but with the same gamma of 2.3.

The colour shifts are now more pronounced than in the previous figures. In the (x, y)-diagrams, the shifts appear to be away from a central region where x and y are roughly comparable. This region includes the white line (not drawn), along which lie the chromaticities of black body radiators. The shifts in both types of diagrams are more pronounced the further the initial colour is from the central region.

Figures 4.13, 4.14, 4.15, and 4.16 present the calculated colour differences when the OSA PlaneSight calibration is applied to a monitor with Conrac monitor phosphor primaries but with the same gamma of 2.3.

The pattern of the colour shifts in the case of the Conrac monitor is somewhat similar to that which occurs when the gamma is reduced, though the pattern is not as regular. However, the shifts are relatively small compared to those for the NTSC monitor.

Figures 4.17, 4.18, 4.19, and 4.20 present the calculated colour differences when calibration-based parameters for a "typical" monitor are applied to a monitor with P22 phosphor primaries.

The pattern of colour changes for the P22 primaries is again loosely towards a central point, though the shifts appear to form a whirlpool pattern due to the change in the monitor phosphors. The shifts are again relatively small compared to those for the NTSC monitor.

In the $L - j = 0$ plane, there are four sample points which fall outside the OSA PlaneSight monitor gamut as the result would require the red phosphor to emit at a negative intensity. Since these negative values were near zero, it was deemed sufficient for the purpose of OSA PlaneSight to force these colours into the gamut by setting the minimum red output to zero. Other solutions to this problem are

briefly discussed in Chapter 6.

Figures 4.21, 4.22, 4.23, 4.24, 4.25, and 4.26 illustrate the shifts in colour caused by changed in illuminant on surface reflectance data taken from several samples of coloured papers, provided by Al Paeth while at the Computer Graphics Laboratory of the University of Waterloo, and from enamel paints[WS82].

The CIE has standarized illuminants for use in colorimetry. The standard illuminants D, such as D_{65}, are intended to simulate natural daylight. The standard illuminant A is realized by a gas-filled coiled-tungsten filament lamp operating at a correlated color temperature of 2856 K. Note that common household incandescent light bulbs also employ a coiled-tungsten filament. We employ these illuminants below. There is currently no CIE standard for fluorescent illuminants. Instead, we employ an approximately fluorescent illuminant based on data given in [WS82].

In the figures discussed below, the calibration white of each light source is represented by a geometric figure. Standard illuminant A is indicated by a square; standard illuminant D_{65}, a circle; a fluorescent illuminant, a diamond; an illuminant with the addition of ambient white light, a triangle.

Figures 4.21 and 4.22 illustrate the shift when the illuminant is changed from CIE standard illuminant A to CIE illuminant D_{65}. The differences in these diagrams are fairly large compared to the differences caused by monitor changes. There are two trends: one is an general shift toward to the left side of the diagram, as the illuminant white shifts; the other is a tendency of the gamut described by the colours to expand.

Figures 4.23 and 4.24 illustrate the shift when the illuminant is changed from CIE standard illuminant A to a fluorescent illuminant[WS82]. The differences in these diagrams are again fairly large compared to the differences in the case of

monitor changes. There are two trends: one is an irregular twisting of the colours about an axis, as the illuminant white shifts; the other is a tendency of the gamut described by the colours to contract.

In figures 4.25 and 4.26, the shifts suggest a convergence on a point which happens to be outside the boundary of real colours. Though the shifts are quite large, the trend is quite regular and predictable.

Figures 4.27 and 4.28 illustrate the shifts in colour when an ambient white light with chromaticity $(0.333, 0.333)$ is introduced to the scene. The colours tend to converge towards a central point corresponding to the white illuminant. The shift is more pronounced for the more saturated colours. It should be noted that high levels of ambient light can restrict the effective gamut of the CRT, as the perceived light becomes more a function of ambient light than monitor output.

In these last four examples of changes in colour due to changes in illuminant, the illuminants are similar to those encountered in everyday life. The human visual system has relatively little problem in compensating for these widely ranging illuminants and perceiving a seemingly full gamut of colour. The shifts caused by changes in monitor parameters are generally quite regular and small in comparison to the shifts due to changes in illumination, yet the human visual system is capable of compensating for such large illuminant shifts. Therefore one can expect that a model-based calibration used to render reflective surfaces will result in correct colour appearance in the presence of common miscalibrations.

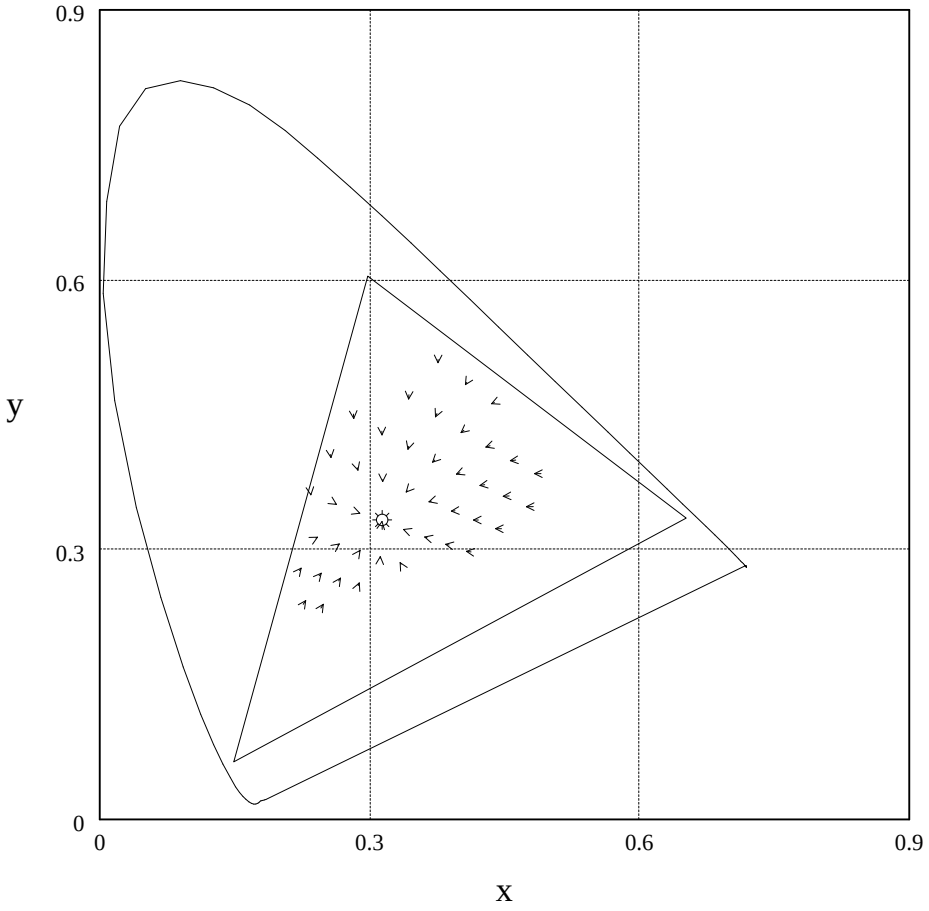

Figure 4.1: CIE (x, y) chromaticity shifts for OSA plane $L = 0$ when the OSA PlaneSight calibration with $\gamma = 2.3$ is applied to a monitor with $\gamma = 2.2$.

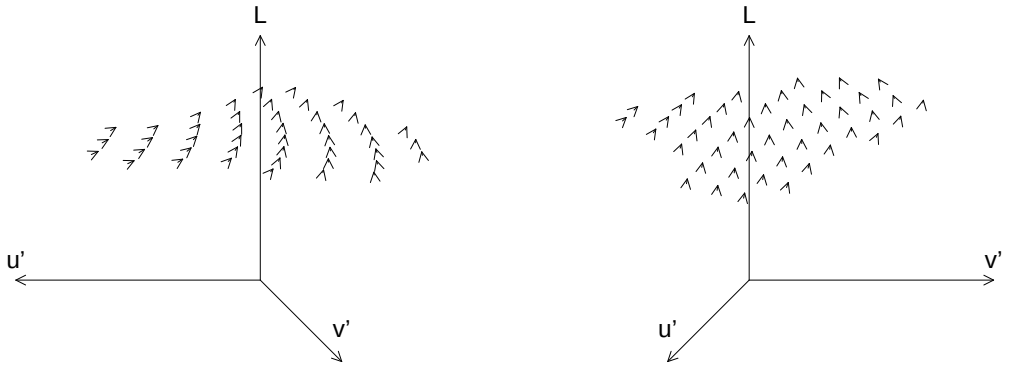

Figure 4.2: Two views of CIE (L^*, u^*, v^*) shifts for OSA plane $L = 0$ when the OSA PlaneSight calibration with $\gamma = 2.3$ is applied to a monitor with $\gamma = 2.2$.

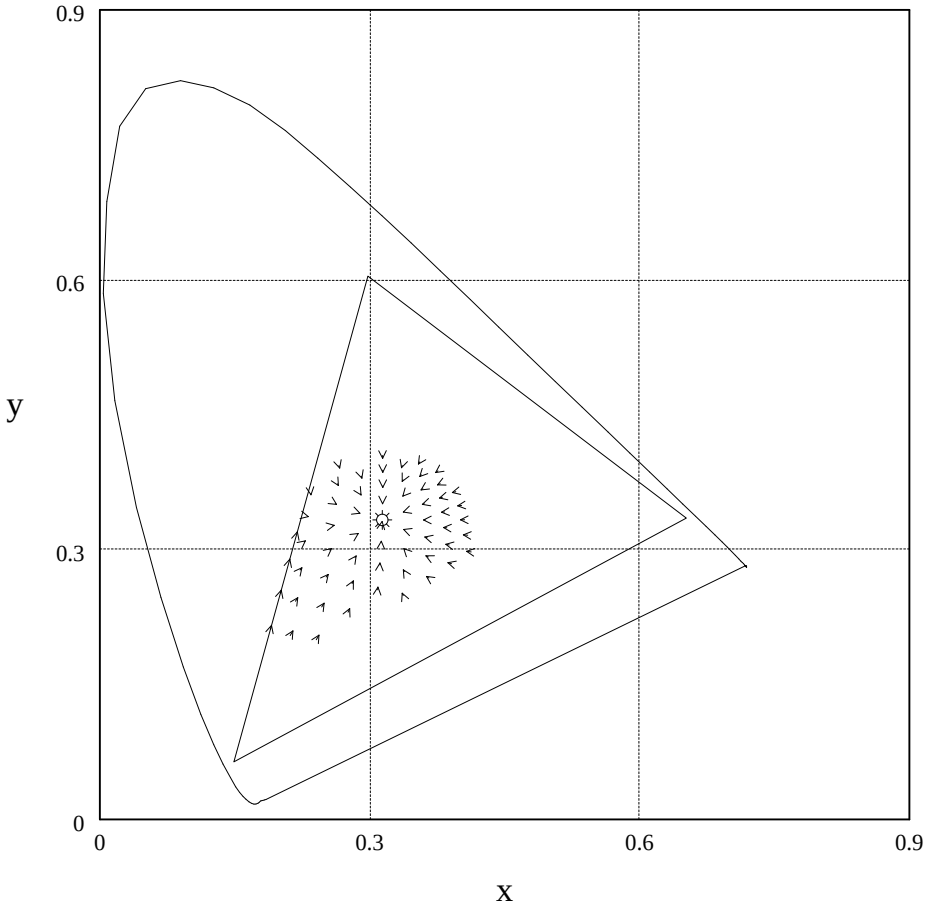

Figure 4.3: CIE (x, y) chromaticity shifts for OSA plane $L - j = 0$ when the OSA PlaneSight calibration with $\gamma = 2.3$ is applied to a monitor with $\gamma = 2.2$.

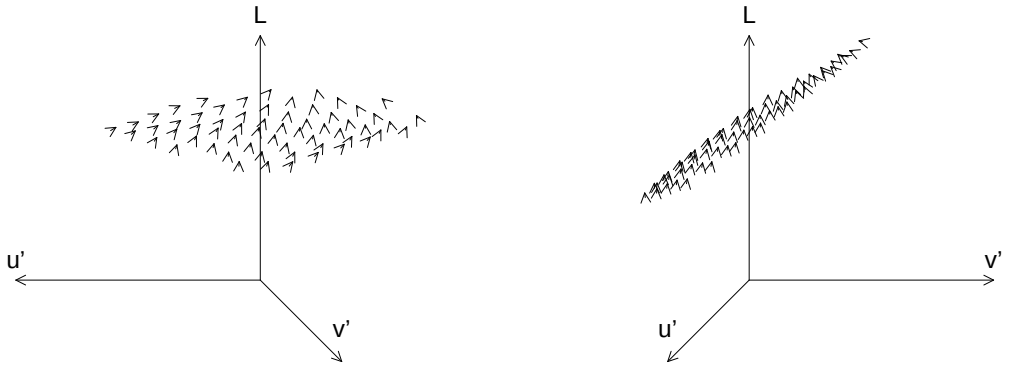

Figure 4.4: Two views of CIE (L^*, u^*, v^*) shifts for OSA plane $L - j = 0$ when the OSA PlaneSight calibration with $\gamma = 2.3$ is applied to a monitor with $\gamma = 2.2$.

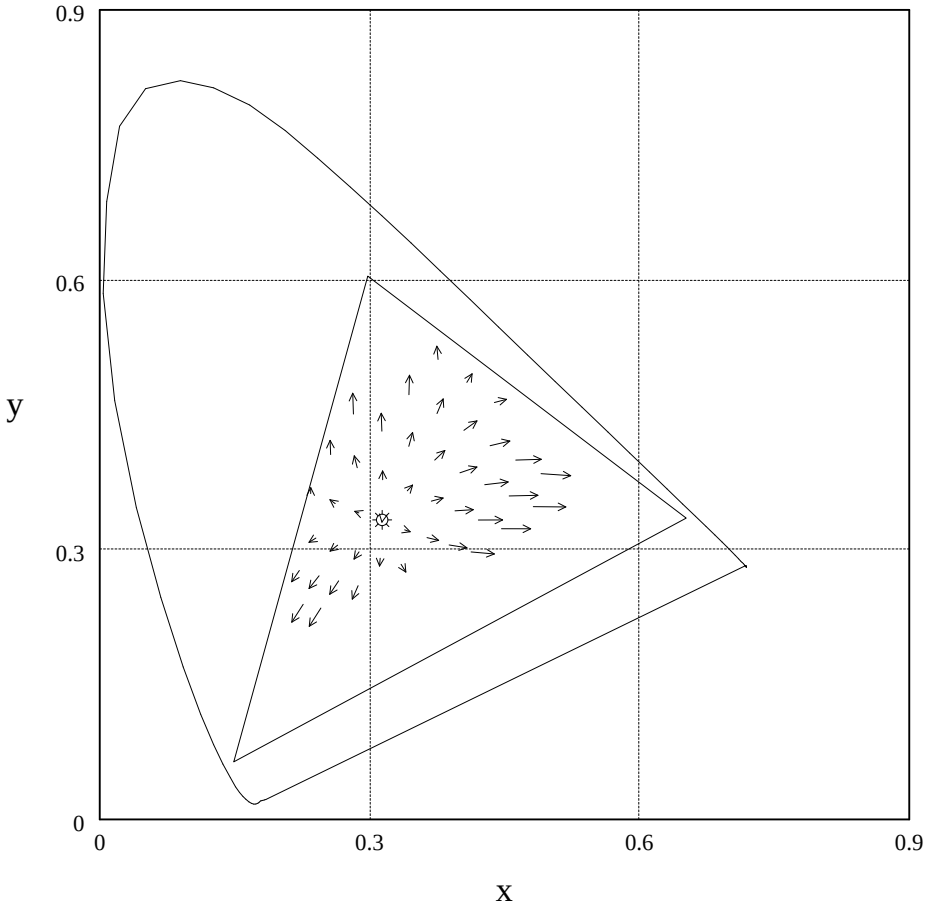

Figure 4.5: CIE (x, y) chromaticity shifts for OSA plane $L = 0$ when the OSA PlaneSight calibration with $\gamma = 2.3$ is applied to a monitor with $\gamma = 2.8$.

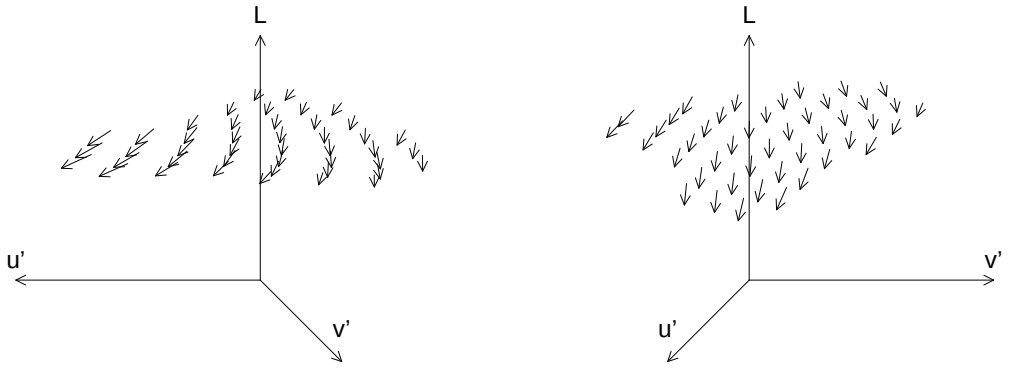

Figure 4.6: Two views of CIE (L^*, u^*, v^*) shifts for OSA plane $L = 0$ when the OSA PlaneSight calibration with $\gamma = 2.3$ is applied to a monitor with $\gamma = 2.8$.

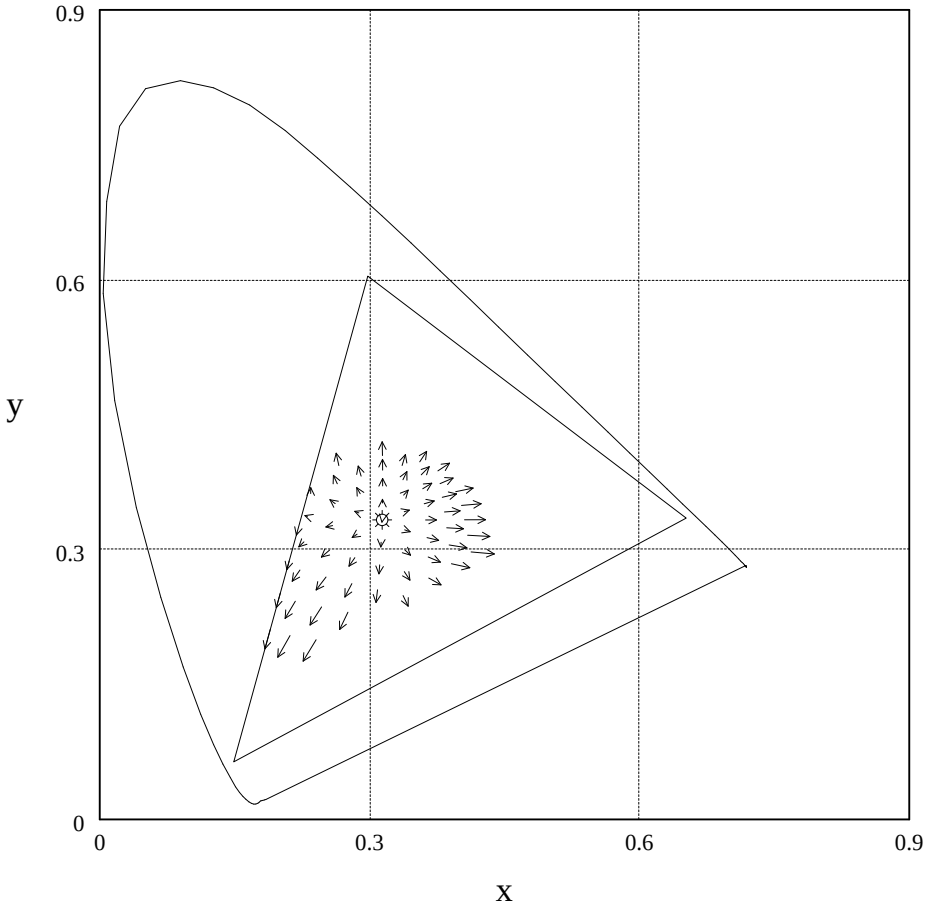

Figure 4.7: CIE (x, y) chromaticity shifts for OSA plane $L - j = 0$ when the OSA PlaneSight calibration with $\gamma = 2.3$ is applied to a monitor with $\gamma = 2.8$.

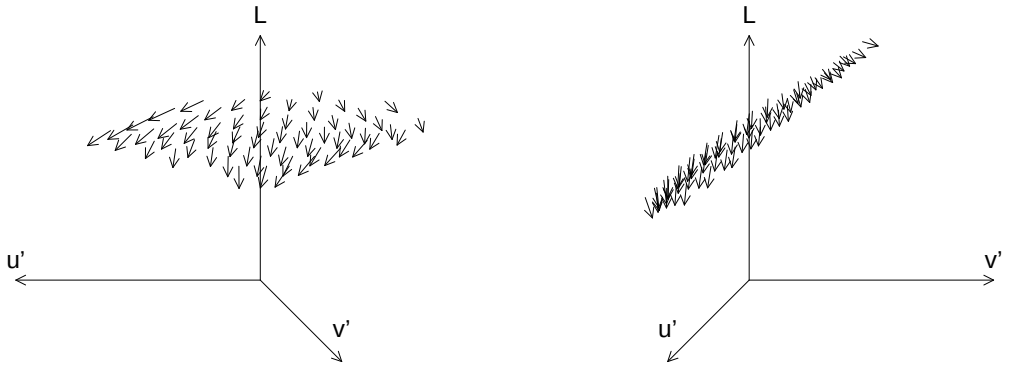

Figure 4.8: Two views of CIE (L^*, u^*, v^*) shifts for OSA plane $L - j = 0$ when the OSA PlaneSight calibration with $\gamma = 2.3$ is applied to a monitor with $\gamma = 2.8$.

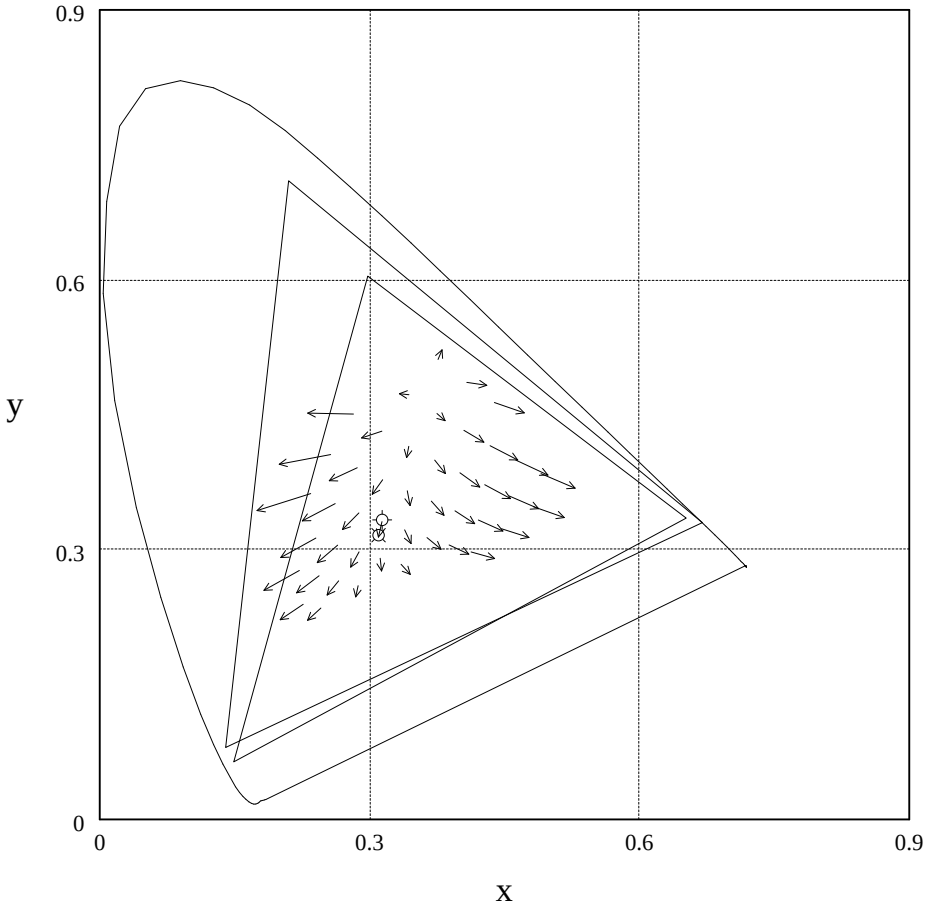

Figure 4.9: CIE (x, y) chromaticity shifts for OSA plane $L = 0$ when the OSA PlaneSight calibration is applied to a monitor with NTSC phosphors.

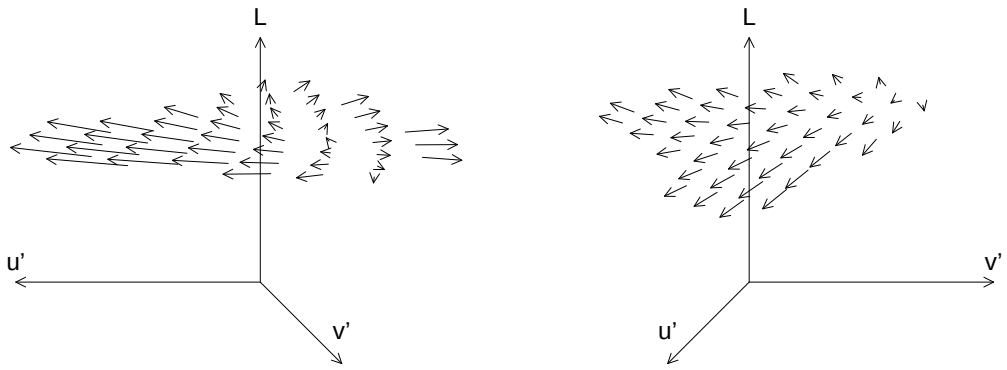

Figure 4.10: Two views of CIE (L^*, u^*, v^*) shifts for OSA plane $L = 0$ when the OSA PlaneSight calibration is applied to a monitor with NTSC phosphors.

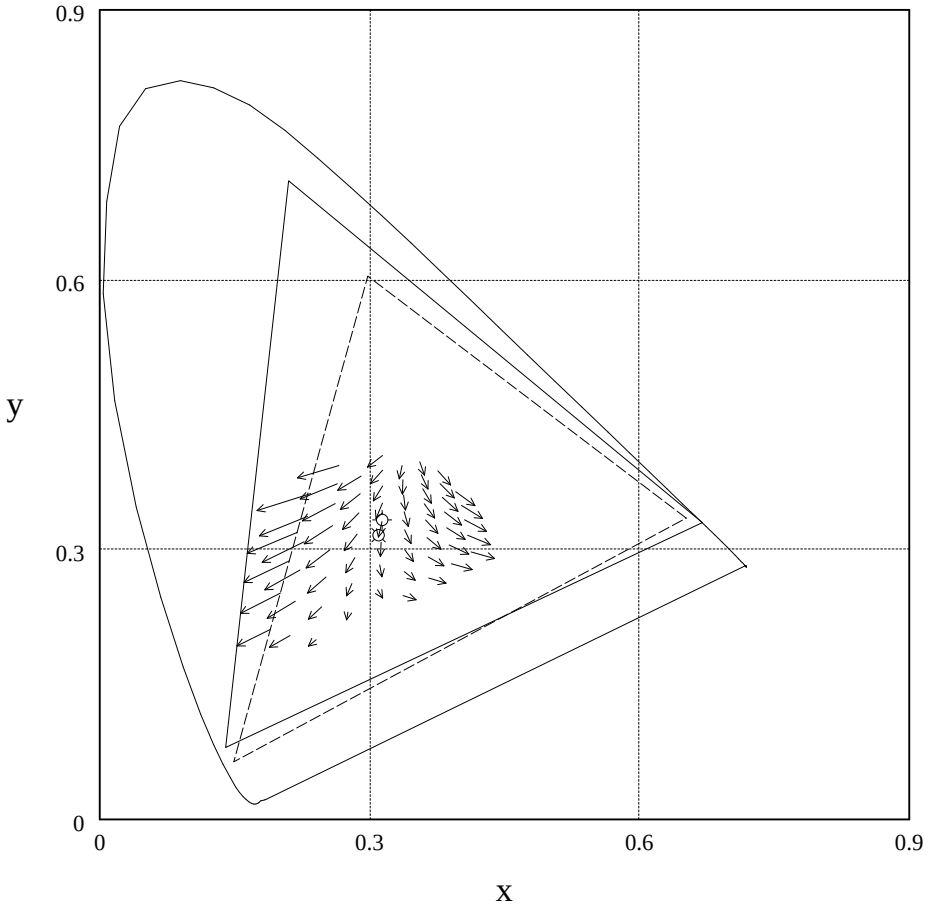

Figure 4.11: CIE (x, y) chromaticity shifts for OSA plane $L - j = 0$ when the OSA PlaneSight calibration is applied to a monitor with NTSC phosphors.

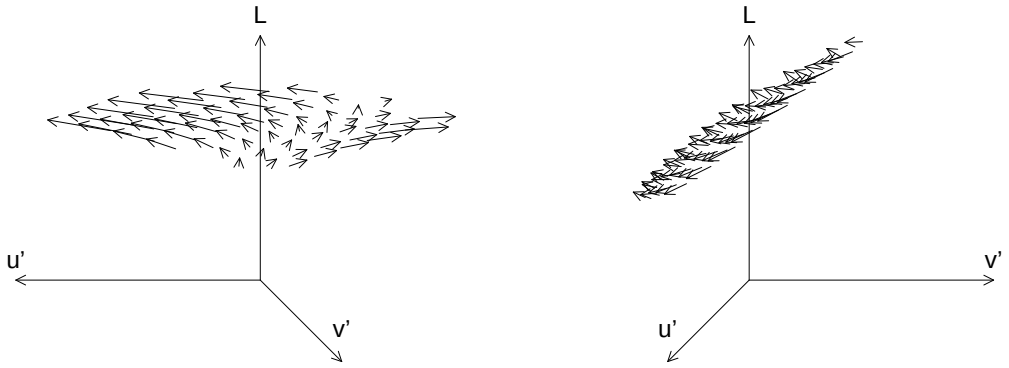

Figure 4.12: Two views of CIE (L^*, u^*, v^*) shifts for OSA plane $L - j = 0$ when the OSA PlaneSight calibration is applied to a monitor with NTSC phosphors.

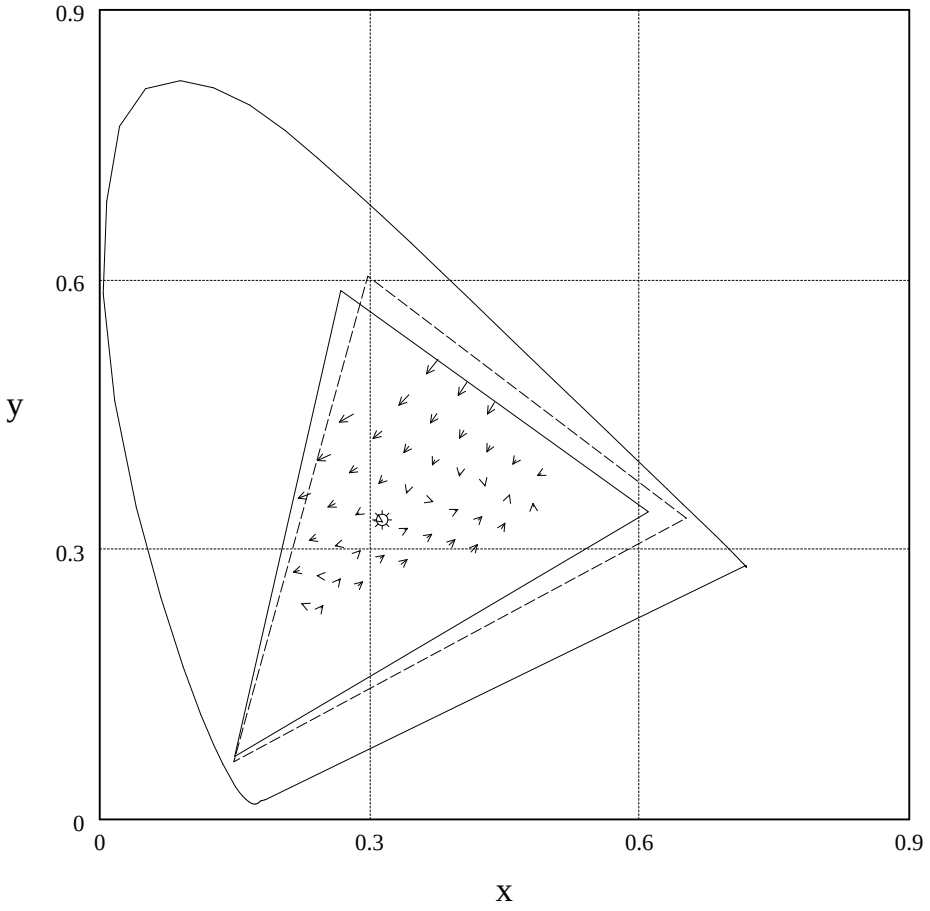

Figure 4.13: CIE (x, y) chromaticity shifts for OSA plane $L = 0$ when the OSA PlaneSight calibration is applied to a monitor with Conrac phosphors.

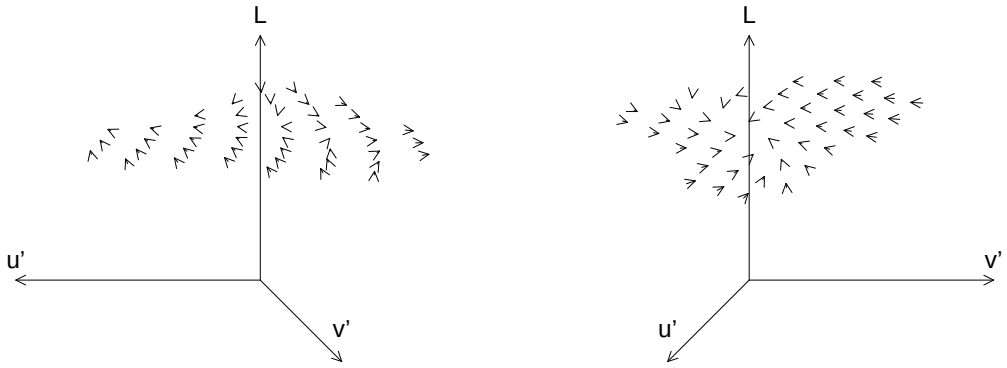

Figure 4.14: Two views of CIE (L^*, u^*, v^*) shifts for OSA plane $L = 0$ when the OSA PlaneSight calibration is applied to a monitor with Conrac phosphors.

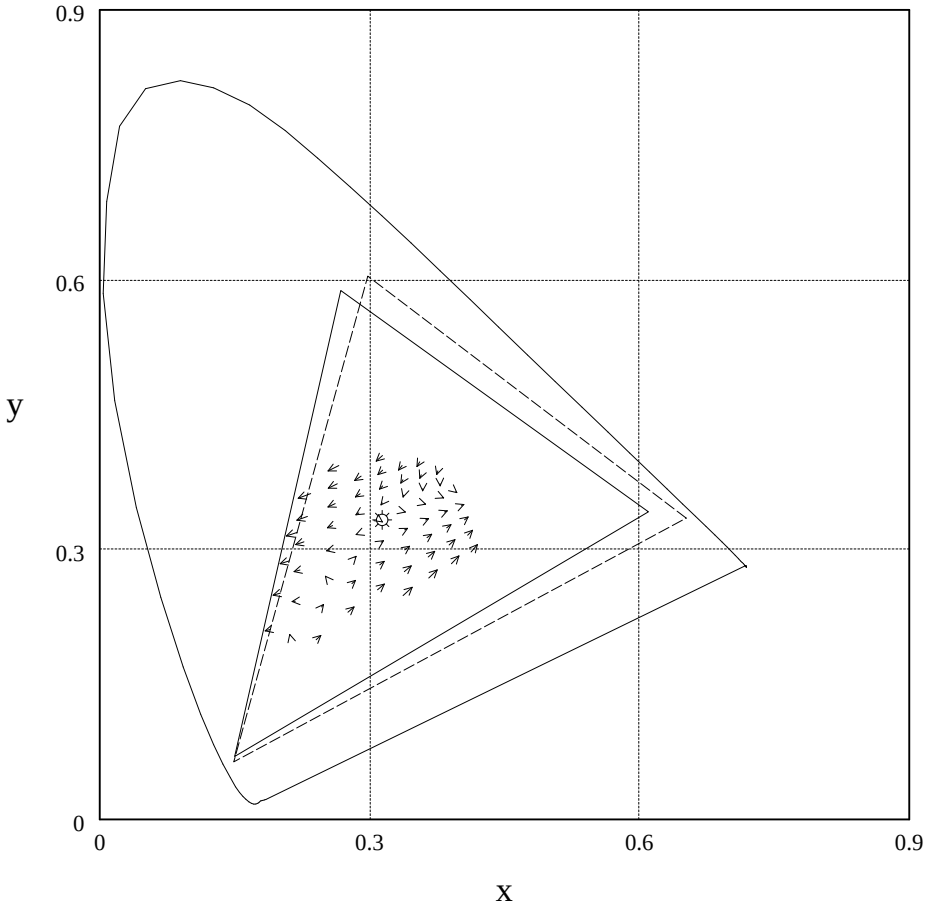

Figure 4.15: CIE (x, y) chromaticity shifts for OSA plane $L - j = 0$ when the OSA PlaneSight calibration is applied to a monitor with Conrac phosphors.

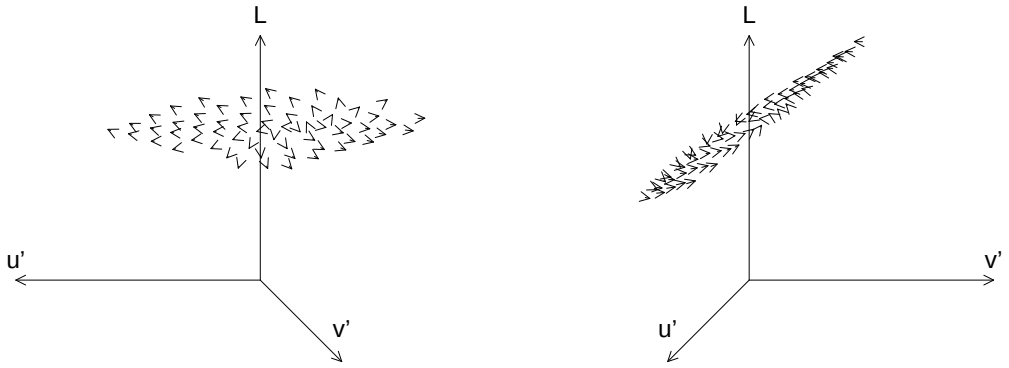

Figure 4.16: Two views of CIE (L^*, u^*, v^*) shifts for OSA plane $L - j = 0$ when the OSA PlaneSight calibration is applied to a monitor with Conrac phosphors.

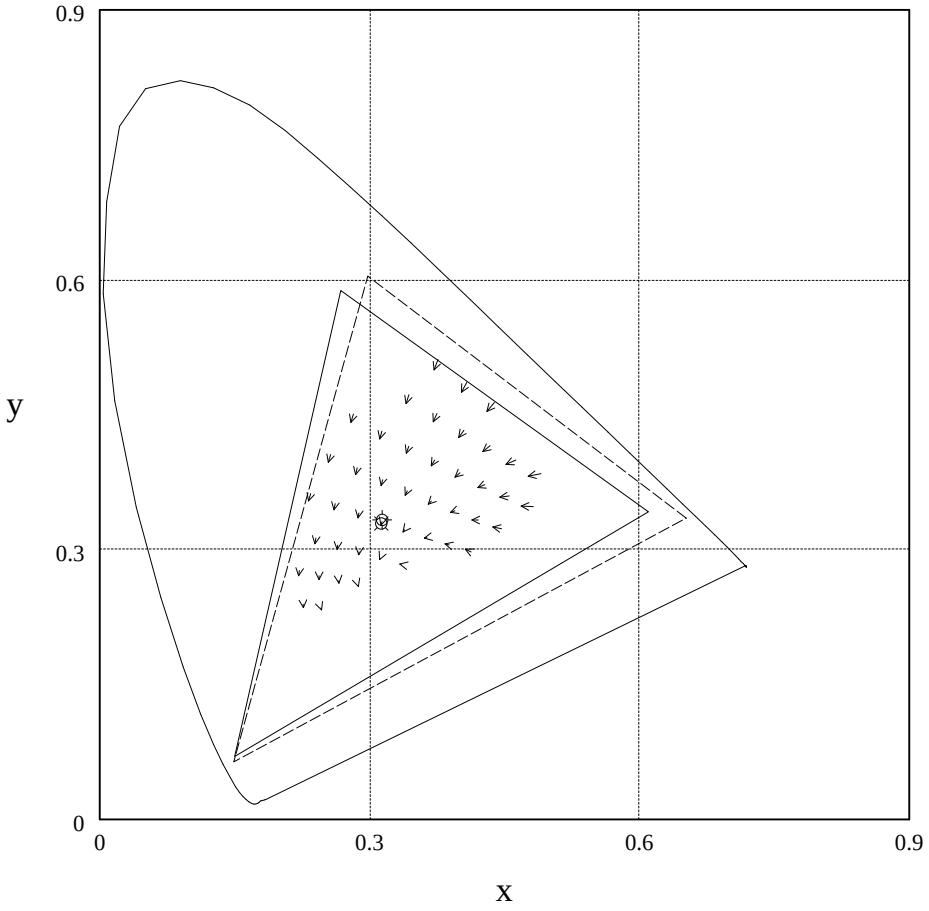

Figure 4.17: CIE (x, y) chromaticity shifts for OSA plane $L = 0$ when the OSA PlaneSight calibration is applied to a monitor with P22 phosphors.

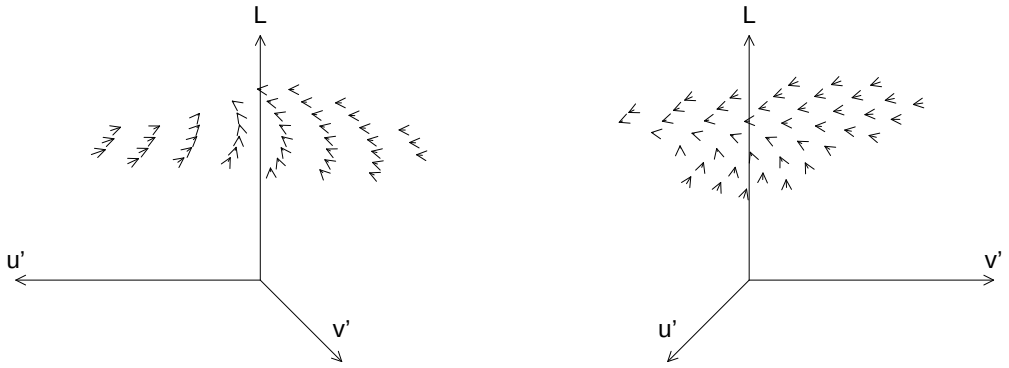

Figure 4.18: Two views of CIE (L^*, u^*, v^*) shifts for OSA plane $L = 0$ when the OSA PlaneSight calibration is applied to a monitor with P22 phosphors.

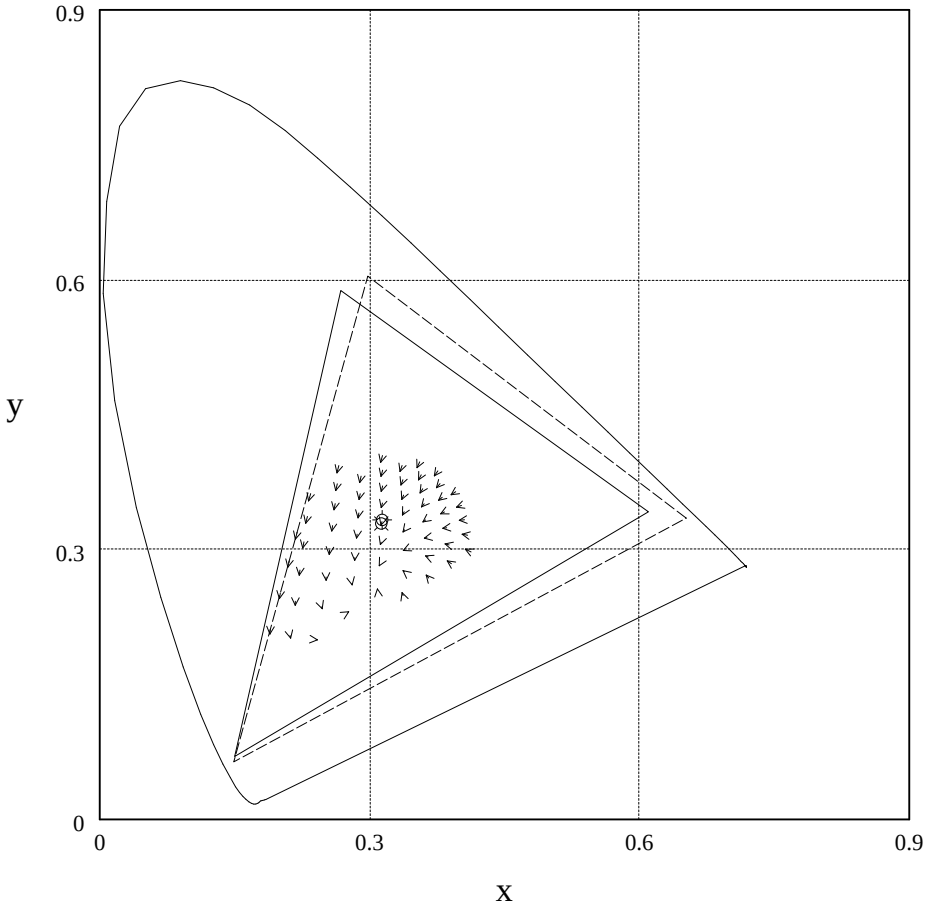

Figure 4.19: CIE (x, y) chromaticity shifts for OSA plane $L - j = 0$ when the OSA PlaneSight calibration is applied to a monitor with P22 phosphors.

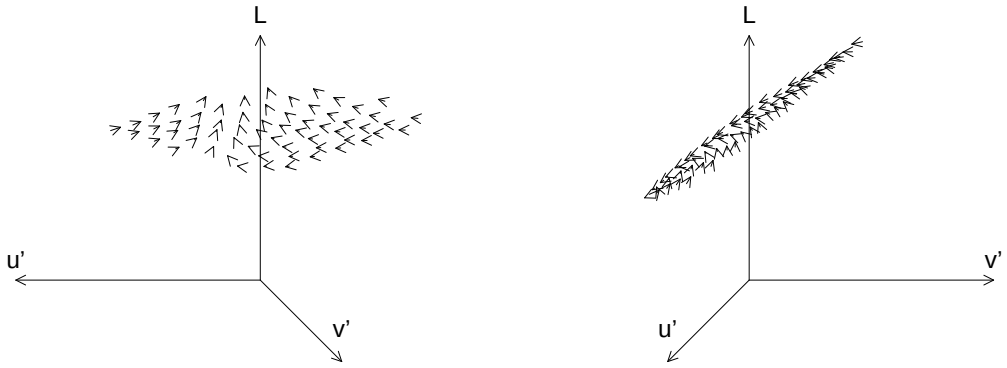

Figure 4.20: Two views of CIE (L^*, u^*, v^*) shifts for OSA plane $L - j = 0$ when the OSA PlaneSight calibration is applied to a monitor with P22 phosphors.

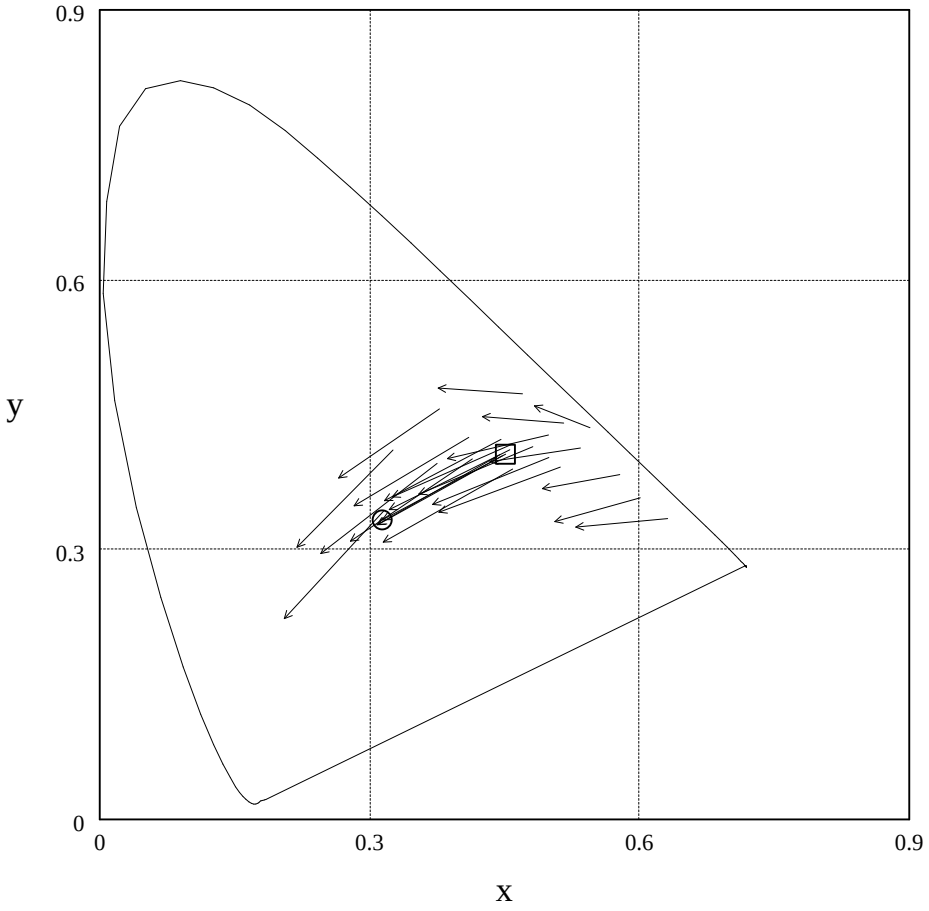

Figure 4.21: CIE (x, y) chromaticity shifts for sample reflectances resulting from illuminant change from CIE A to CIE D_{65}.

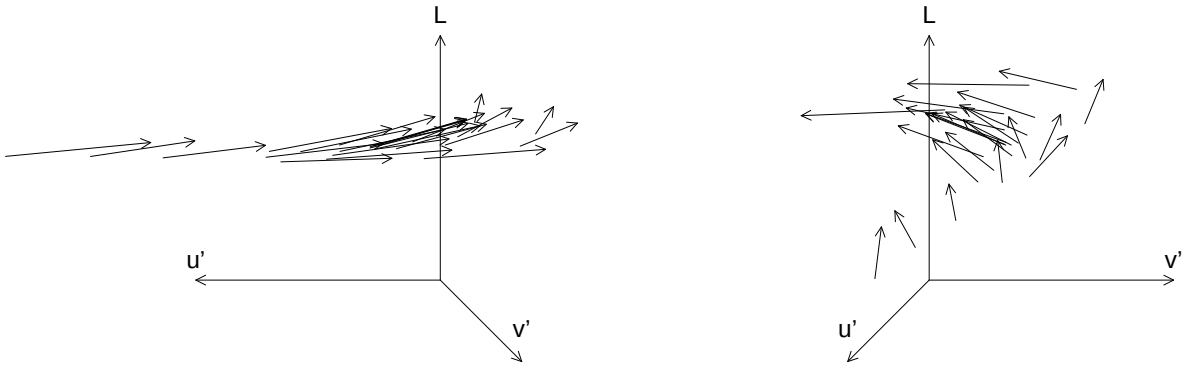

Figure 4.22: Two views CIE (L^*, u^*, v^*) shifts for sample reflectances resulting from illuminant change from CIE A to CIE D_{65}.

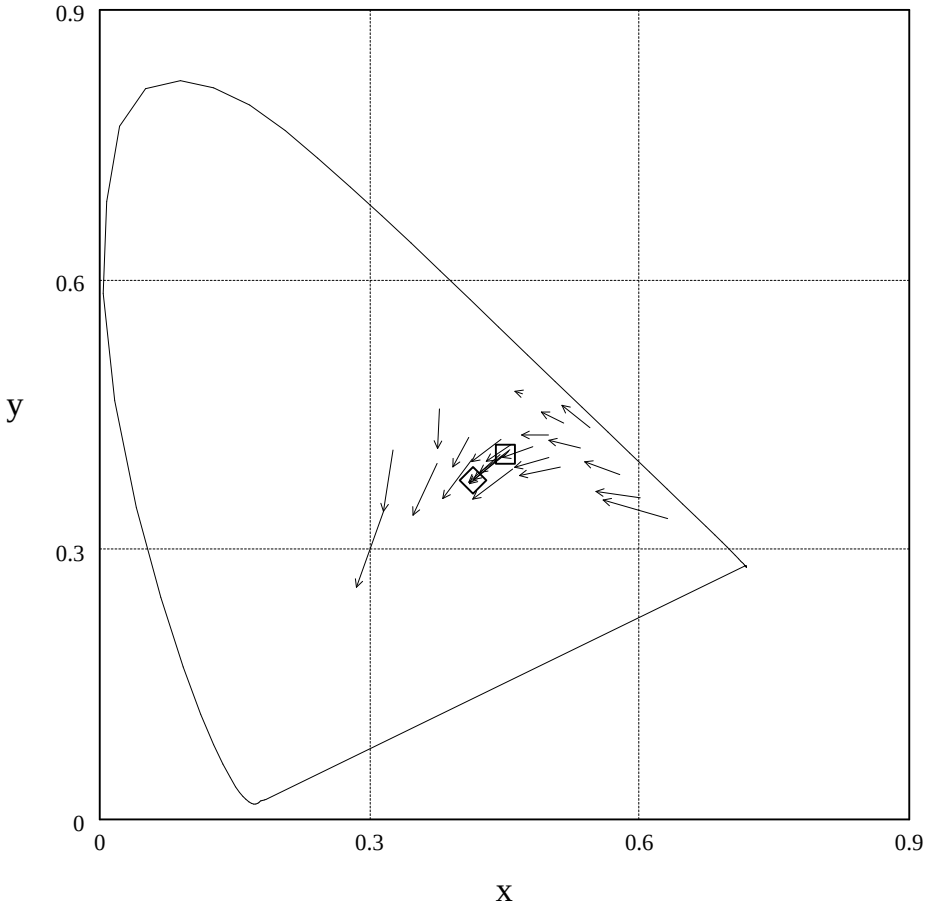

Figure 4.23: CIE (x, y) chromaticity shifts for sample reflectances resulting from change from CIE illuminant A to a fluorescent illuminant.

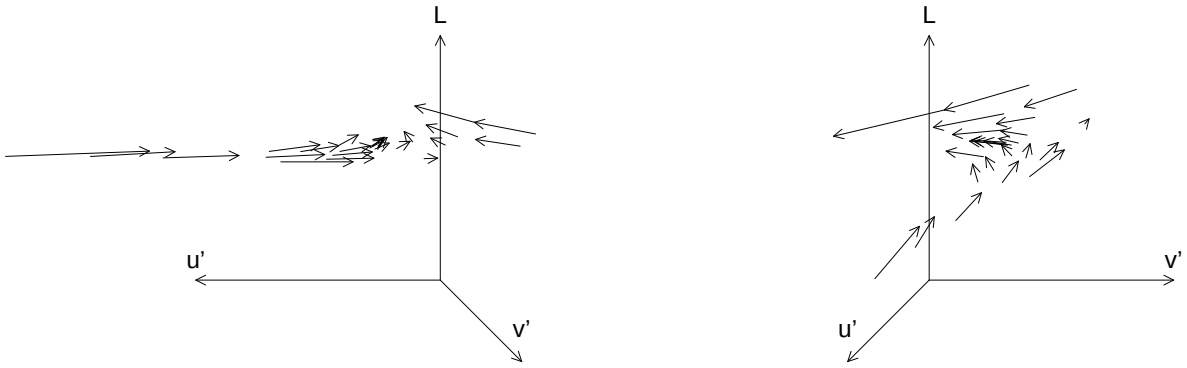

Figure 4.24: Two views of CIE (L^*, u^*, v^*) shifts for sample reflectances resulting from change from CIE illuminant A to a fluorescent illuminant.

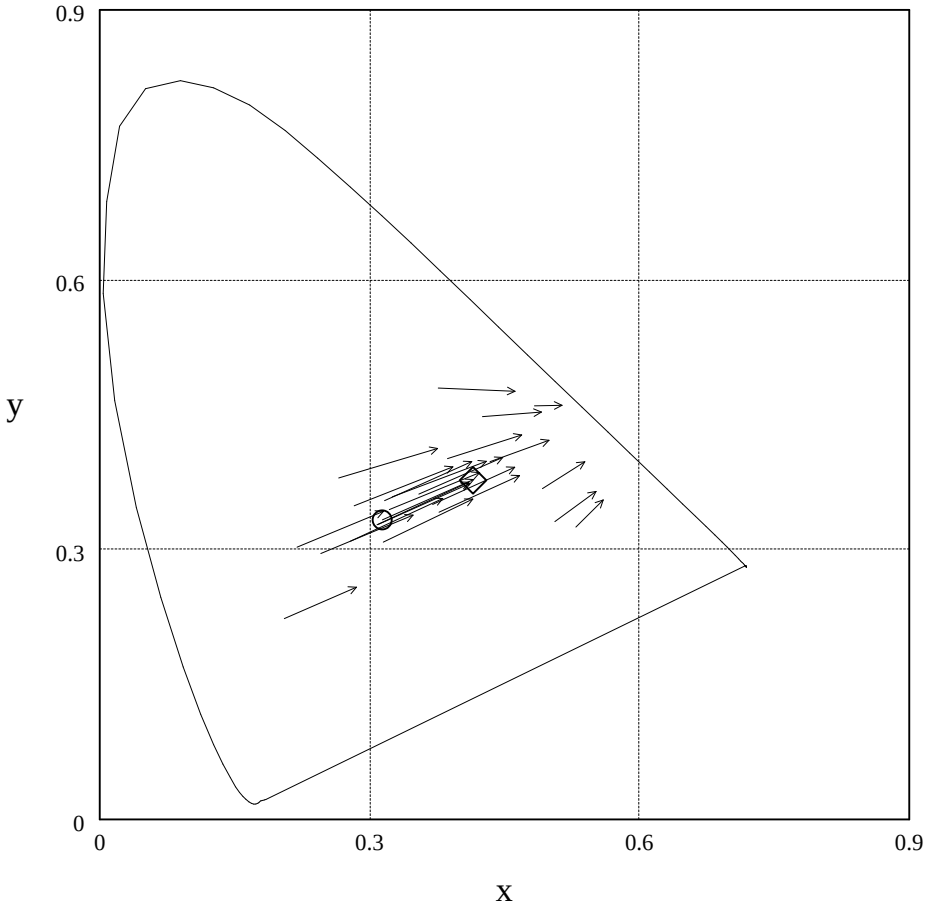

Figure 4.25: CIE (x, y) chromaticity shifts for sample reflectances resulting from change from CIE illuminant D_{65} to a fluorescent illuminant.

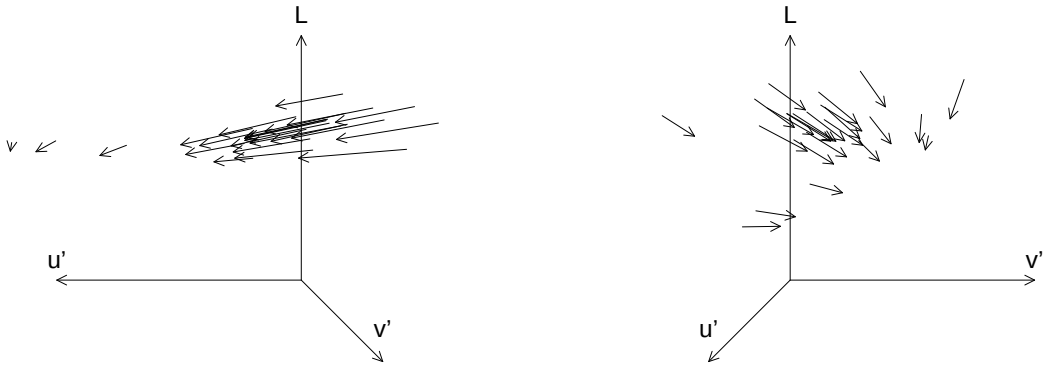

Figure 4.26: Two views of CIE (L^*, u^*, v^*) shifts for sample reflectances resulting from change from CIE illuminant D_{65} to a fluorescent illuminant.

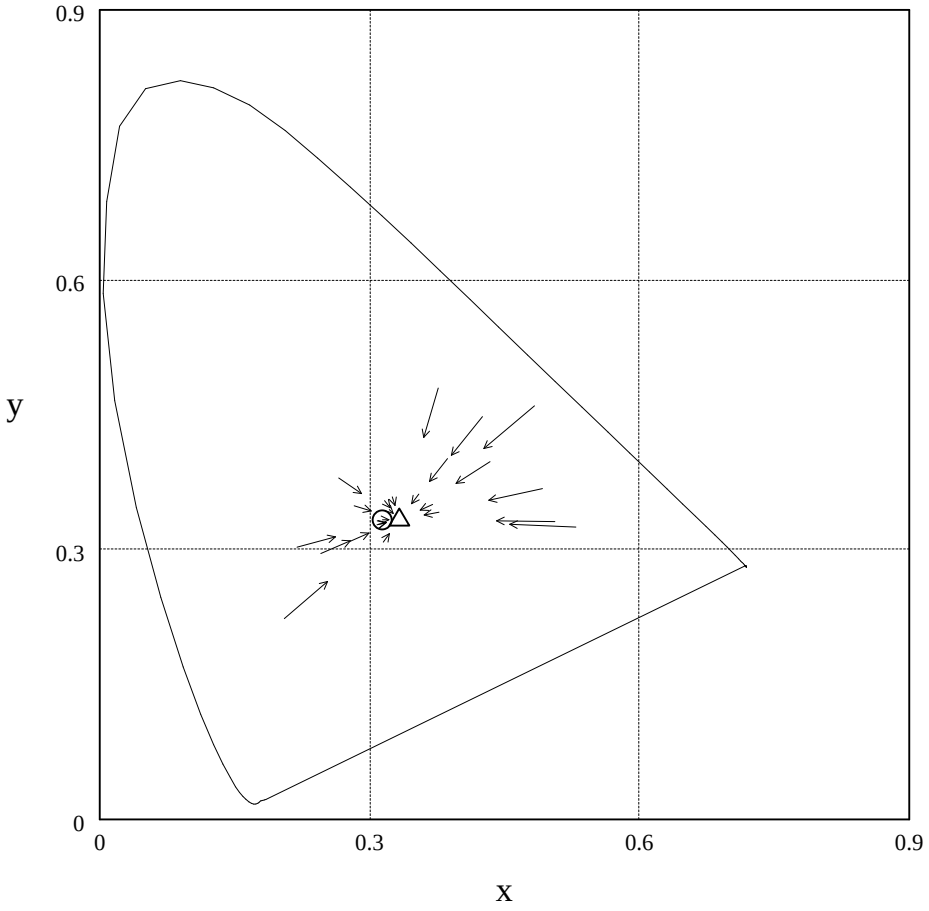

Figure 4.27: CIE (x, y) chromaticity shifts for sample reflectances resulting from illuminant change from CIE D_{65} caused by the addition of white light.

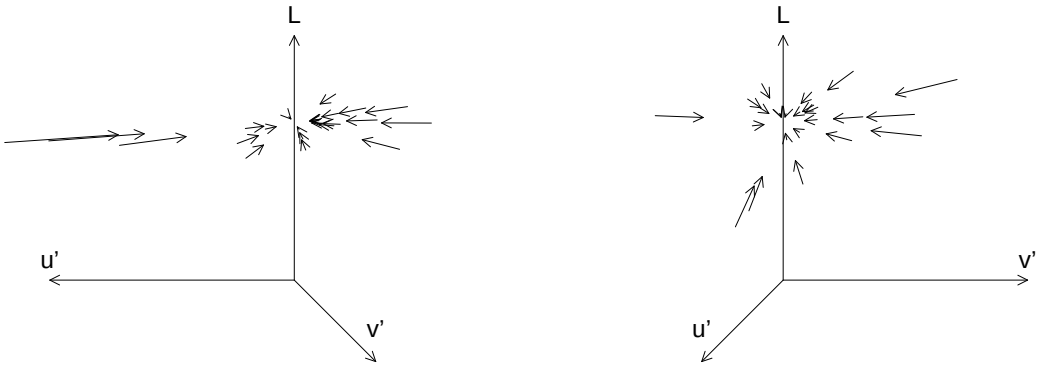

Figure 4.28: Two views of CIE (L^*, u^*, v^*) shifts for sample reflectances resulting from illuminant change from CIE D_{65} caused by the addition of white light.

Chapter 5

Group Theory and Interface Design

5.1 Introduction

This chapter discusses the construction of the interface for OSA PlaneSight. A Macintosh computer served as a platform for the development of OSA PlaneSight. Since the Macintosh interface is an example of good design, the basic guidelines behind its construction are presented briefly and used in evaluating the OSA PlaneSight interface. The geometry of the OSA space is presented briefly, this time emphasizing underlying symmetries that can be exploited by a navigation model. Group theory is introduced as a general tool for constructing compact user interactions. An interaction for OSA PlaneSight is then derived using group theory. The initial implementation of the user interface is discussed, followed by response and criticism by both casual users and user interface experts. A revised interface is then presented.

5.2 The Macintosh User Interface Guidelines

The success of the Macintosh owes much to the general consistency and intuitiveness of its interface. As OSA PlaneSight is implemented on a Macintosh, it is important that its interface conforms with the Macintosh interface guidelines.

The philosophy behind the Macintosh interface is that it should appeal to an audience which does not program, and may even be distrustful of computers. To this end, the manufacturer's guidelines[App85] suggest that all programs, or applications, on the Macintosh should embody the qualities of *responsiveness*, *permissiveness*, and *consistency*. These attributes are discussed briefly below.

Responsiveness, simply stated, means that an action made by the user should have a direct, immediate result. Actions should be intuitive, and thus quickly learned.

Permissiveness means that the user should be allowed to do anything reasonable. In a program, the term *mode* refers to a program state that the user explicitly enters and leaves. Modes restrict the operations available to the user while in effect. Permissiveness requires the avoidance of modes except in unusual cases when they are the natural or the best way to perform an operation.

A large set of user interface routines provided by the Macintosh performs several basic functions. Consistency of the interface design allows both novice and experienced users to take advantage of previous experience with the interface when faced with new or different applications. Programmers are not restricted to using only the existing routines, but new features should be designed so as not to conflict with existing conventions.

5.3 The Interaction Paradigm

Here we briefly cover the geometry of the OSA Colour System, which was discussed in far more detail in Chapter 3. We then discuss the viewing paradigm used in navigating the space and some design decisions implemented in OSA PlaneSight.

The sample points of the OSA Colour System are placed in a regular rhombohedral lattice of equidistantly spaced points in three-dimensional space. Each point in this lattice is immediately surrounded by twelve equidistant points. Through each point there are three square planes, oriented at right angles to each other, and four triangular planes, oriented relative to each other like the faces of a tetrahedron. The square planes are specified by the plane equations $L = const$, $g + j = const$, and $g - j = const$; while the triangular planes are specified by $L + g = const$, $L - g = const$, $L + j = const$, and $L - j = const$. For the plane to be nonempty, the plane must intersect the gamut of the OSA space and the constant in the plane equation must be an integer. Furthermore, the constant must be even in the case of the triangular planes. The geometry is more extensively explained in Section 3.8 of Chapter 3.

The three coordinates (L, j, g) form a right-handed coordinate system for the OSA space. However, there is a more natural coordinate basis for the OSA space. Let us define the coordinate basis (b_1, b_2, b_3) to be $b_1 = 2L$, $b_2 = g + j$, $b_3 = g - j$. The square planes are then specified by the plane equations $b_1 = const$, $b_2 = const$, $b_3 = const$. The triangular planes are specified by $a-b-c = const$, $a+b+c = const$, $a-b+c = const$, and $a+b-c = const$. Thus, this basis better reflects the symmetry inherent in the arrangement of the OSA space. For this reason, this basis is used when rendering the OSA space.

OSA PlaneSight defines a viewing reference frame in which the viewing orien-

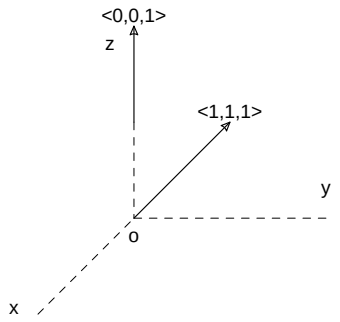

Figure 5.1: Viewing axes and vectors of OSA PlaneSight

tation of the selected OSA plane is defined. Let us define the axes of the viewing reference frame to be: x, y, and z. Initially, the axes of this reference frame are set to b_2, b_3, and b_1 respectively. In this frame, looking at the OSA space down the $(0, 0, 1)$ direction allows one to view square planes of colours; while looking down the $(1, 1, 1)$ direction allows one to view triangular planes of colours.

The method of navigation between planes is constrained in OSA PlaneSight. When moving between parallel planes, the choice is restricted to to one of the two nearest parallel planes, subject to the constraint that the destination plane must contain at least one OSA colour. Movement between nonparallel planes is done by selecting a point that is to lie on the desired plane and then rotating the viewing axes with respect to the axes of the OSA space. These movements are described in more detail below and in Section 5.5.

One of two operations is required to move between parallel planes, depending on the direction of movement. One operation increases the constant in the plane equation, and one decreases the constant. The proper increment for moving to the nearest parallel plane for the square planes $L = const$ is 1, while the proper increment for the other planes is 2. It is straightforward to determine these increments

from the plane equations and the constraints on the position of the sample points.

The rotation operations R_x and R_y perform 90° rotations of the OSA space about the x- and y-axes of the viewing reference frame respectively, and are sufficient to reach all possible permutations of axes in a right-handed coordinate system. The rotation operation R_z, which performs a 90° rotation of the OSA space about the z-axis, was added for symmetry and convenience; R_z can be composed from the sequence: R_x, R_y, R_x, R_x, R_x. Only one additional operation is required to select which of the two viewing vectors $(0, 0, 1)$ and $(1, 1, 1)$ is being used.

There are problems with combining these two methods of maneuvering. Because of the way sample points are positioned in the OSA space, a former center of rotation after being translated to the nearest parallel plane does not lie on a valid colour. A rotation about that point can result in a plane containing no OSA colours, particularly if the resulting plane is oblique. This artifact can be seen in the plane equation restrictions. Another possibility is that the formerly valid rotation point after being translated to a parallel plane now lies outside the OSA Space gamut. After a translation has been performed in OSA PlaneSight, any selected rotation point is discarded. Consequently, the user is required to select a new center of rotation. In keeping with the Macintosh interface, the point is selected with the mouse.

Another problem involves the nature of translation and rotation, and the modelling of the effects of a rotation. It is best demonstrated by example. Let us refer to the center of rotation as a pivot point. Because rotation is not commutative, the displayable colours may slowly "crawl" out of bounds of the display window. When the display position of pivot point is preserved through each rotation, chosing a different pivot point between each rotation allows one to translate the colours. One solution is not to preserve the display position of the pivot point during a rotation.

The result is described better as selecting between pages of an album containing the various defined OSA planes.

The initial pivot point in OSA PlaneSight was chosen to be the 30% reflectance grey, which lies at the origin of the OSA coordinate system. The initial plane orientation was arbitrarily chosen to be the constant lightness plane $L = 0$.

5.4 Group Theory in a Nutshell

A method of modeling user interactions using group theory is proposed below. To this end, the concept of a group must first be formally defined. A group is defined to be a set G with an operator $*$ that satisfies the following axioms.

1. G is closed with respect to $*$.
2. $*$ is associative.
3. There exists an element e in G such that $a * e = a$ and $e * a = a$ for every element a in G. The element e is the identity.
4. For every element a in G, there exists an element a^{-1} in G such that $a*a^{-1} = e$ and $a^{-1}*a = e$. Thus every element has an inverse with respect to the identity element.

The symbol $\mathbb{Z}$ denotes the set $\{..., -3, -2, -1, 0, 1, 2, 3, ...\}$. Some of the simplest finite groups consist of integers modulo n, denoted by the symbol $\mathbb{Z}_n$, with the operator being addition modulo n, where n is an arbitrary positive integer.

The direct product of any two groups G and H, denoted by $G \times H$, is also a group and is defined on a set of ordered pairs:

$G \times H = \{(x, y) : x \in G \text{ and } y \in H\}$.

If the operators for G and H are denoted as $*_G$ and $*_H$, respectively, then the operator $*$ for $G \times H$ is defined as follows.

$(x_1, y_1) * (x_2, y_2) = (x_1 *_G x_2, y_1 *_H y_2)$.

5.5 Group Theory and the Interface

As we shall see below, the application of group theory can provide a compact description for user-interaction, and assist in the derivation of an interaction that is complete, consistent, permissive, and minimal. In short, by using group theory as the basis for a model of user-interaction, simple and intuitive user interfaces can be derived.

Before group theory can be applied, the user-interaction must be stated formally. Let S denote a set of states in an interaction and F denote a set of operations, where operations are functions defined by the state transitions. An operation f which models an aspect of the user interaction can be defined as a partial function $f : S \mapsto S$, $f \in F$.

The general model of operations stated above is not useful in determining what operations are appropriate, as there are no restrictions on what one can do. Too much generality allows the creation of arbitrary, nonintuitive operations.

A useful restriction is to model the user-interaction by a group S' defined on the set S. Operators can be defined on group S' as partial functions $f_i(x) = x * y_i$ where $x \in S$ and $y_i \in S$; x corresponds to the present state and the y_i entries determine the mapping from x to $f_i(x)$. Because of the restrictive nature of groups, it is important to determine what information should be represented by each state and what group

operator should be chosen. For well-known groups and Cartesian products of such groups, the choice of group operator is straightforward.

As an example of modelling a user interaction with a group, recall that the implemented interface allows a user to rotate from one plane orientation to another and to switch viewpoints. A possible choice of group is $S = \mathbb{Z}_2 \times \mathbb{Z}_4 \times \mathbb{Z}_3 \times \mathbb{Z}_2$. Recall that the group $\mathbb{Z}_n$ of integers modulo n has the operator of addition modulo n. The first group $\mathbb{Z}_2$ is used to indicate whether or not the colour space is rotated 180° about the x-axis before viewing. The group $\mathbb{Z}_4$ corresponds to which of the four rotations of the colour space about the viewing coordinate z is selected, either $0°, 90°, 180,°, 270°$. Together, $\mathbb{Z}_2 \times \mathbb{Z}_4$ determine which octant of the colour space is in the first octant of the viewing reference frame. $\mathbb{Z}_3$ then determines which of the three cyclic permutation of the axes of the colour space is in effect. $\mathbb{Z}_2 \times \mathbb{Z}_4 \times \mathbb{Z}_3$ thus defines the possible orientations of the colour space with respect to the viewing reference frame. The last group $\mathbb{Z}_2$ determines which of the two viewing directions is in use. The group $\mathbb{Z}_2 \times \mathbb{Z}_4 \times \mathbb{Z}_3 \times \mathbb{Z}_2$ defines 48 states, which results in 2304 possible state transitions and 48^{48} possible operations. Clearly, only a small fraction of the possible operations are worth considering.

As part of the implementation of OSA PlaneSight, an attempt was made to choose an appropriate group that yielded a complete, consistent, compact, and permissive interaction. Ideally one constructs a complete set of operations, that is, a set of reasonably-intuitive operations that are general enough to reach any desired state from any other state, possibly invoking a large number of operations. In the above example, observe that an approach using group theory encourages the division of states into orthogonal attributes; an operation that takes advantage of such division should be more consistent and more intuitive than an arbitrary operation. Another consequence of division into orthogonal attributes is that redundancy can

be spotted and eliminated, thus yielding a more compact interface. Operators that perform 180° and −90° rotations are redundant as they can be composed of successive 90° rotations. However, not all redundancy should necessarily be eliminated. For instance, the R_z operation can be accomplished using a sequence of R_x and R_y operations. Because the R_z operation is symmetric with the R_x and R_y operations, its exclusion is arbitrary and unintuitive. Thus, a permissive interface includes the R_z operation to improve the ease of use of the interface.

The group given above as an example is not used in OSA PlaneSight because the operations thus modelled are not intuitive and do not communicate the symmetry of the OSA color system well to an inexperienced user. A simple solution involves modelling the rotations using the group $M \times \mathbb{Z}_2$, where M is a group of 3×3 matrices generated by the rotation operations R_x and R_y. (Recall that a rotation in n dimensions can be represented by an $n \times n$ matrix.) This latter approach was implemented.

Yet another advantage of group theory is the compactness of the resulting notation. For instance, the control which toggles which of the two viewing vectors is to be used can be described by the function $f(x) = x * (I, 1)$ over the set $M \times \mathbb{Z}_2$. Moving between parallel planes can be modeled using the group $S = \mathbb{Z}_n$, where the two controls correspond to $f_1(x) = x + 1$ and $f_2(x) = x + (n - 1)$. (Note that the addition of $n - 1$ modulo n is the inverse of addition 1 modulo n.) Operations which can be described simply are likely candidates for intuitive operations. It was found that the wrap-around nature of the movement between parallel planes was non-intuitive, so the actual implementation redefined $f_1(n - 1) = n - 1$ and $f_2(0) = 0$. In this case, the interaction derived by group theory was more powerful than was required. Nevertheless, the use of group theory served as a means of verifying that the interaction was not unduly redundant or confusing.

A minor weakness of using group theory is that there is no convenient way for an operation to set an attribute directly within the restrictions on group operators. Such operations can serve useful purposes, such as resetting state variables to an initial state or to preferred states. For example, a switch modelled with $\mathbb{Z}_2$ can be toggled, but not set on or off regardless of the current state. A more general operator that directly sets a single attribute can be used to more accurately model such an interaction.

In summary, group theory provides a compact description of a user interaction. It also assists in the derivation of a complete, consistent, compact, and permissive interaction. Because the derived group theory models of the OSA PlaneSight interactions are compact, the code to implement them was straightforward to write. The goals of a minimal interface and a permissive interface can be contradictory at times, so an acceptable compromise may be necessary. It is generally not desirable to allow the user to use all possible distinct operations. However, the applied restrictions should not appear to limit unduly the freedom of the user to invoke desired operations. The allowed state transitions should be carefully examined. It is possible that a large number of state transitions are not required in an interaction. Reducing the number of operations leads to a reduction in the number of controls needed for an interaction. A large number of controls tends to increase the difficulty of learning the interaction for the first time, possibly making the interaction awkward. In addition, each control takes up space, which is at a premium on a system with a window-based interface.

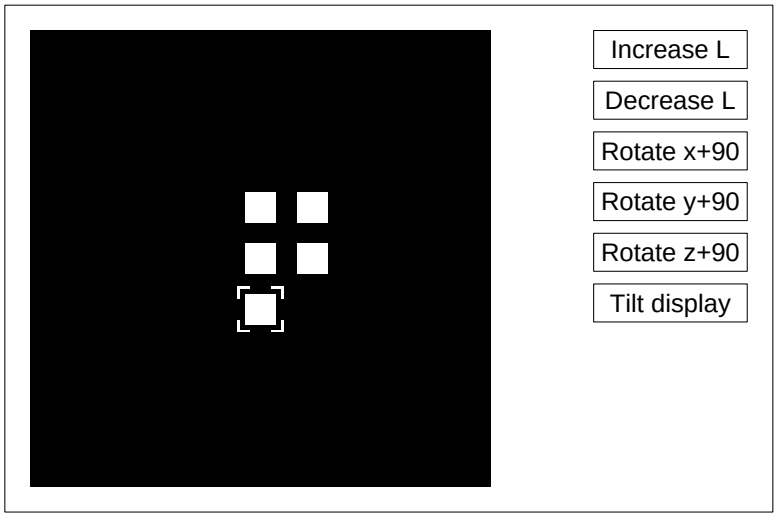

Figure 5.2: Schematic layout of OSA PlaneSight interface

5.6 Initial Implementation of the Interface

Though OSA PlaneSight was implemented on a Macintosh, the application could have been implemented on many architectures. At the time the decision was made, the Macintosh provided what seemed to be the simplest architecture available. A more detailed description of the architecture and implementation is presented in Appendix A. The initial implementation described below was later revised based on user response.

The Macintosh employs an overlapping-window-based graphical user-interface. By default, the root window covers the screen. Applications, including the operating system, interact with the user via windows, which are displayed as if they overlap the root window. The order in which these other windows superpose each other is not fixed and is often changed by the user. Windows provided by the system are generally rectangular, although other window shapes can be created.

OSA PlaneSight occupies a single window on the screen; this is roughly depicted

in Figure 5.2. The majority of the window is occupied by a black background square, flush to the left. The OSA colours are displayed on this square. The plane is actually displayed rotated 45° so the squares surrounding the colours line up with the square pixels of the monitor. See Appendix A for a more detailed discussion of this design choice. The colours are separated from each other by borders of the background, rather than being adjacent. To the right are several control buttons which allow the user to navigate through the space. The colour selection palette is also placed here in the revised design. To activate a button, the user moves the mouse cursor on top of the control button and presses the mouse button down. The resulting plane is then displayed within the background square.

Six buttons are used in the interaction. The first two are used to step between parallel planes. If there is no immediately parallel plane in a given direction, the corresponding button is "greyed-out" to indicate such action is not possible. Stepping between parallel planes is done by incrementing or decrementing the constant in the appropriate plane equation. The next three buttons are used to select x-, y-, and z-rotations of the viewing axes; the viewing axes are described in Section 5.3. The final button toggles the current viewing vector between the two viewing vectors as defined earlier.

For the square lattice planes, it is natural to choose squares for the shape of the colour samples being displayed. The colour samples appear as squares rotated 45 degrees counterclockwise from the natural screen coordinates, if j, g is directly mapped to a Cartesian x, y. Other shapes for the colour samples were considered and rejected. The problem with other shapes is that curves and slanted lines give the appearance of jagged edges due to rasterization, as the resolution of the monitor is limited. This has been traditionally solved by the use of antialiasing. However, antialiasing requires additional colour determination. The additional colours so

rendered will change under varying monitor calibrations, possibly being noticeably different from the intended colour. The extent of this problem was not fully investigated. An implementation of antialiasing also requires the writing of routines to support the drawing of such regions, erasing such regions, and determining if a point is inside these rotated squares. To avoid these complications, the viewing reference frame was rotated by 45 degrees when rendering the colours on the screen. The triangular planes are rendered using square colour samples. This has the benefit of preserving the shape and size of the individual colour samples in all plane orientations.

5.7 Casual User Response

Regardless of the theory behind the design of an interface, the only way to determine its effectiveness is to test it with a group of users. The users who tested OSA PlaneSight consist of members of the Computer Graphics Lab at the University of Waterloo, a few computer science students not studying graphics, and several employees of the Human Interface Group of Apple Corporation at Cupertino, California, including artists and interface designers. Some of the recommendations received are discussed below, as are subsequent modifications to OSA PlaneSight. The interest in OSA PlaneSight was quite positive as quite a number of constructive comments and suggestions were received.

There was some debate over the best shape with which to display the colours. Squares are used in the rendering of the OSA space for both rectangular and triangular planes in the initial design. Consistency of the interface is furthered by choosing the same shape for both types of planes. A natural shape for the triangular planes is the hexagon. However, hexagons do not give the appearance of a

uniform arrangement when laid out in a rectangular grid. Circles were also suggested, but were not implemented because it was felt that the limited resolution of the display made rendering aesthetically pleasing circles difficult, if not impossible without antialiasing. In addition, the user-interface routines provided with the Macintosh include a procedure to determine if a point is enclosed within a rectangle. Testing if a point is enclosed within a hexagon or circle requires that new functions be written.

One suggestion made several times, in slightly different forms, was for some form of three-dimensional navigational aid to help users to visualize the OSA colour space. This aid would help inexperienced users learn to navigate through the space. One person suggested a geometric illustration, using coloured spheres to represent the relative arrangement of the plane currently being displayed and how it related to other planes. Another suggested that some form of animation be used to illustrate the rotation of the viewpoint. Yet another suggested a display of the (parallel) planes which one could step to or the planes which one can rotate to, allowing the viewing neighboring colours.

Interestingly, observation of a few casual users revealed that they were able to develop a feel for maneuvering about the various OSA planes after only a few minutes of practice. One strategy developed for selecting colour was to step through the planes of one orientation until a desired colour was found and then examine the colours on the other planes passing through that colour. However, experience indicates that users need to be encouraged to spend the time needed to familiarize themselves with the program.

There was a proposal that planes other than those defined could be displayed. Unlike other discrete colour spaces used in art and design, colours not explicitly specified in the OSA Colour System can be readily determined via the equations

defining the space. It is possible that interpolation methods may offer a good approximation of such colours implicit in the OSA space, i.e. enclosed by the defined gamut but not listed in the OSA specification. This was not investigated as it is beyond the scope of OSA PlaneSight.

There were a few objections to using a black background in the initial design, as opposed to the 30% reflectance grey commonly thought in the graphics industry to be the most neutral of colours. The black background had been chosen because it would add to the apparent brightness of the colours in an environment with high ambient illumination. Because the OSA colours were originally calibrated against a 30% reflective background, use of a black background in OSA PlaneSight reduces the perceived different between simple colors. In addition, the difference between the brighter of the colours in the OSA space and the black background appeared rather jarring due to the high contrast. Consequently, the program was revised to use a 30% grey background.

Another suggestion that was implemented was to provide a means of selecting multiple colours for a colour scheme. It was suggested that there be a means of judging colour choices that is dependent on the relative areas that the colours would occupy on a screen, thus allowing for colour context. In the revision, we added a set of rectangles that colours can be assigned to and recalled from; these rectangles are located at the lower-right corner of the window. A surrounding colour provides a background for the seven other colours. One of these colours occupies a larger rectangle than the others since it is intended to serve as a foreground colour. The current colour chosen as a rotation point is indicated by a surrounding white crosshair, while other selected colours which happen to lie on the plane being viewed are indicated by a surrounding black crosshair.

There are many methods of assigning colour to and retrieving colour from a

palette, and no agreement on which is best. The method used in the revised implementation is to toggle between assignment and retrieval mode. In assignment mode, a colour is assigned to a palette entry by using the mouse to select a color from the displayed OSA plane and then selecting the palette rectangle to which the colour is to be assigned. When in retrieval mode, a colour is retrieved from the palette by selecting a palette rectangle with the mouse. Any translation needed to reach the plane containing that colour are performed; the current viewing orientation is otherwise preserved. There are other possibilities of course. One alternative include providing separate assignment and retrieval controls, which would require a means of indicating palette entry selection. Another is for colours to be selected and "dragged" to and from the palette; this would require a means of differentiating between selection for palette entries and for points of rotation.

The changes implemented in the revised version of OSA PlaneSight are generally sound, although a few aspects have not been properly evaluated. There has not yet been sufficient opportunity to gather user feedback on their efficacy and each of use.

Chapter 6

Conclusion

6.1 Evaluation

The application developed, called OSA PlaneSight, allows a user to maneuver through the uniform colour space defined by the OSA Colour System; this application has been implemented on a Macintosh computer. By using group theory as a tool for analysis, a simple, complete, consistent, permissive, and compact user interaction model for navigation was derived, as well as an interaction model for the selection of a colour scheme. It is straightforward to implement a similar tool on any computer system that is capable of displaying a wide range of colour using a CRT monitor. The interface as currently implemented is easily learned, albeit the modifications made subsequent to user response have not been properly tested.

Group theory has been presented as a tool in the construction of interaction paradigms. This approach has the advantage of encouraging the division of the relevant states according to orthogonal attributes, and helping to eliminate unwanted redundancy. The notation provided by group theory itself is sometimes insuffi-

cient to express commonly desired operations conveniently, such as an operation to set an attribute explicitly. This minor problem can be resolved by generalizing to functions that can directly assign values to variables, in addition to functions that invoke group operators.

An error-tolerant approximation to a calibrated monitor for non-critical applications was achieved by using a gamma correction model to characterize the monitor. This allows an approximation to a uniform colour space to be presented in a prototype tool for colour selection that can be employed in a wide range of viewing conditions.

A tool similar to OSA PlaneSight can be used as a basis for a tool capable of resolving conflicts between the aesthetic and the functional usage of colour in a user-interface by incorporating heuristic-driven construction of colour schemes. Such tools would be an significant advance in the colour management for applications such as the window-based workstation environment. A better understanding of how context affects human perception of colour and shape, and thus information transmitted, is required before such a tool can be perfected.

Tools that provide access to a uniform colour space have a wide range of practical applications. For instance, in scientific visualization, when colour is used to represent a change in a quantity, it is advantageous to have the perceived change in colour be linearly related to the change in that quantity. Such a tool could be used to select the endpoints of a gradation to be used in a scalar range. It was beyond the scope of this thesis to determine whether or not a linear interpolation of the endpoints is an acceptable technique for generating such a gradation.

6.2 Open Questions

The minimal interaction model presented may not be the most intuitive or convenient interaction model possible. Whether a better interaction model exists is an open question.

Extension of the OSA PlaneSight gamut used in colour selection to all theoretically possible OSA colours to the entire monitor gamut ended in failure. The gamut of the OSA Colour System sample points does not encompass many monitor colours that are often used, highly saturated colours in particular. Many of these do not correspond to reflectances of stable paints, and in some instances do not correspond to reflectances achievable by real objects.

It became clear that the curvature of the OSA space far outside the bounds of the defined sample points makes the space inadequate for representing colour out to the boundary of the monitor gamut. This curvature causes the space to exhibit instabilities, which makes calculation of the inverse tedious by the method described and quite possibly erroneous perceptually.

A uniform colour space based on a photoreceptor model, as opposed to a cube-root formula that is the basis of the OSA space, offers a possible solution to this problem. The SVF colour space[SV86] may be better suited for the problem of gamut extension, although the formulae are not easily extended to low luminosities.

Another problem in extrapolating the colour model to the full monitor gamut is that monitor bounds vary with the different phosphors. This means that applications requiring high accuracy need separate calibrations for each monitor. For non-critical applications, it may be that a "typical" or compromise monitor calibration may suffice.

The problem of OSA colours falling inside or outside the monitor gamut depend-

ing on the phosphors can be remedied in several ways. The solution implemented is to force the colours just outside the gamut to the boundary of the gamut. Another solution is to discard the out-of-gamut colours. A third solution, which may be the best, is the use a virtual set of monitor phosphors as the calibration monitor. The colours can then be guaranteed to lie in the gamut, and the presentation made acceptable. This was not done in OSA PlaneSight since the validity of colour constancy as applied to monitor rendering was not properly established at the time.

There was insufficient time to implement heuristics for generating colour schemes. Integrating such heuristics with a colour selection tool similar to OSA PlaneSight may provide a means of partially automating colour scheme selection; the tool would generate one or more colour schemes based on the selection of a small number of colours by the user.

Appendix A

Implementation Details

Some of the material in this appendix is included from previous chapters to improve clarity.

A.1 The Macintosh Environment

Relevant aspects of the Macintosh architecture are presented below in detail sufficient to allow straightforward construction of similar applications on other architectures[App85, App88].

The Macintosh employs a window-based graphical user-interface. By default, the root window covers the screen. Applications, including the operating system, interact with the user via windows, which are displayed as if they overlap the root window. The order in which these other windows superpose each other is not fixed and is often changed by the user. Windows provided by the system are generally rectangular, though other shapes can be created.

A set of library routines allows applications, or user programs, to interact with

the operating system in a uniform manner. These routines are collectively termed the Macintosh User Interface Toolbox. The motivation behind this toolbox is to provide a means of achieving a degree of hardware independence and user interface consistency. All text and graphics are rendered on a medium resolution bitmapped display directly or indirectly via QuickDraw, a part of the toolbox.

OSA PlaneSight was developed on a Macintosh IIx computer running System 6.0.3, with a RasterOps 24-bit colour board. The program used QuickDraw to access the RasterOps hardware. Bypassing QuickDraw and accessing the hardware more directly would run the risk of making the application incompatible with other graphics boards. Furthermore, the documentation accompanying the RasterOps 24-bit graphics board is insufficient to bypass QuickDraw. Conforming to the QuickDraw interface also has the advantage that the application was able to run under MultiFinder, an application which allows the user to perform multiprogramming.

On the Macintosh, a colour look-up table (CLUT) is associated with each monitor. Each colour table entry includes the 24-bit monitor RGB specification and several attribute flags used by the colour manager. When a colour has to be displayed on a monitor, the colour manager takes the RGB specification and chooses what it believes to be the closest colour table entry available.

Monitor RGB colours can be accessed indirectly in QuickDraw by specifying an index in an application palette. The Color Manager takes the specification in the palette entry and compares it to the CLUT to find the the appropriate RGB signals to send to the monitor frame buffer.

The palette manager attempts to ensure good colour matching, mediates requests for displayed colours when a desired colour is not available, and attempts to avoid colour conflicts with other applications. Colour on a multi-window display is

a shared resource, managed by the palette manager in conjunction with the CLUT. When different application windows are to use different colours, the problem of colour contention arises. To resolve this, each application can specify to the palette manager an intended usage for each entry in its colour palette. Each time a new window is brought to the "front" or an appropriate QuickDraw routine is called, the palette manager determines which window has priority over colour usage and changes the palette colours and usage values accordingly.

Colours can be specified as *courteous*, *tolerant*, *animating*, or *explicit*. Courteous colours are dependent on QuickDraw for appropriate values to be set. Tolerant colours are approximated by the nearest available colour in the CLUT when rendered. Animating colours are reserved and unavailable to other colour requests. Explicit colours ignore such restrictions, forcing the corresponding entry in the CLUT to an explicit RBG value; in general, explicit colours should not be used by application in a multi-application environment.

There is insufficient colour resolution when using the 8-bit colour board to allow each of the OSA colour samples to be distinguished from each other, due to the spacing of the entries in the default CLUT. Eight bits of resolution are sufficient if the colours are chosen carefully and assigned to the CLUT on the video card which controls the monitor (and reassigned as the plane being displayed is changed). This was not pursued further, as the hardware available did not require such strategies to achieve a sufficient number of colours. A palette of sufficient size to encompass the full-step OSA colours was used.

Interactive applications on the Macintosh are driven by events. In general, such applications are structured around an *event loop handler*. The latter is a loop that reads one event at a time from an event queue, then generates a response based on that event and the current state of the program. User input in the form of mouse,

keyboard, and keypad actions, are processed by the Toolbox Event Manager and stored on the *event queue*, which is a FIFO queue. If the number of events to be stored exceeds the capacity of the event queue, the oldest events are discarded. When taking an event off the queue, one can restrict the type of event selected. The event queue can carry other system events, such as activate and update events, that respectively signal that windows are to be activated, deactivated, or redrawn. Additionally, an application can define up to four event types of its own. The QuickDraw drawing primitives do not generate update events.

Two types of mouse events are defined by the Event Manager. Mouse-down events are generated by pressing the mouse button down; mouse-up events are generated by releasing the mouse button. The position of the mouse cursor as displayed on the screen is also stored in these events. A mouse click refers to a sequence of pressing and releasing the mouse button. Because input events have precedence over update events, interactions that require multiple mouse clicks are possible.

The window occupied by the application is to be redrawn only when an update event occurs. It is up to the application to acknowledge the update event and redraw appropriate parts of the window. Activation events can likewise be acted upon or ignored by the application.

A.2 OSA PlaneSight

OSA PlaneSight assumes the monitor has "typical" phosphor chromaticities as given in Chapter 4. Computing the transformation from XYZ to monitor RGB as needed by each plane, without storing it for later use, degraded the response time of OSA PlaneSight until it was essentially unusable interactively, as the computation

involves extensive floating point calculations. The XYZ coordinates of the 424 OSA full-step colours were precomputed and stored in a file. The file is read in at initialization; the XYZ values are then converted to monitor RGB coordinates and then stored in a colour palette.

We judged that a file containing XYZ data was best for portability between applications, since the program does not have to be recompiled if the data is to be changed. If this file is to be updated according to the redetermination of the OSA colours[Mac90], it is straightforward to implement this by recomputing the colour specifications offline and replacing the data file.

The time taken to initialize the palette remains unacceptably long if this application is to be used in conjunction with other applications interactively. Speed can be gained by compiling monitor RGB coordinates for the OSA colours into the program, at the expense of possible recompilation each time the data is changed. A precise calibration for each monitor requires a new version to be compiled; however, we argue in Chapter 4 that such precise calibration is not required for common applications.

Each time OSA PlaneSight receives an update event, it redisplays the controls and draws the currently selected plane and any cursors indicating selected colours on the affected portions of the screen. Judicious use of two procedures, referred to as ValidRect() and InvalidRect(), allows clipping regions to be specified on the screen within which the update redrawing is performed. These regions need not be connected. Refreshing the screen in 24-bit colour mode is sufficiently slow that optimizing these update regions significantly improves performance.

When the user steps through several planes in rapid succession, the program is unable to update the screen until the last plane has been reached, as mouse and keyboard events have higher priority than the update event. This behavior

can be changed by adding code which forces the program to redraw the screen immediately once a new plane is selected. This was not done, as it makes stepping through planes tedious; redrawing the colours in the current plane on the screen is already noticeably slow.

The OSA coordinates are directly indexed for efficient access into a three-dimensional array containing colour palette indices. Most of the entries are empty due to the OSA full-step set construction.

OSA PlaneSight occupies a single window on the screen. The majority of the window is occupied by a background square drawn at 30% grey. The OSA colours are displayed against this background, separated by several pixels of background. To the right are control buttons that allow the user to navigate through the space, and the colour selection palette. To activate a button, the user moves the mouse cursor on top of the control button and depresses it.

It is natural to choose squares as the shape of the colour samples being displayed for the square lattice planes of the OSA space. A thorough discussion of the geometry of the OSA space is given in Chapter 3. Since we are using the viewing reference frame (x, y, z) given in Chapter 5, it would seem natural to align the (x, y) of the frame with the screen (x, y). However, the colour samples then would be rendered as squares rotated 45 degrees counterclockwise from the natural screen coordinates. Because of the limitations of QuickDraw and the resolution of the monitor, there are jagged edges due to rasterization when squares aligned with the pixel arrangement are used. Antialiasing requires additional colour determination, and may degrade under varying monitor configurations sufficiently to be noticeably different from the intended colour. An implementation of antialiasing also requires the writing of routines to support the drawing of such regions, erasing such regions, and determining if a point is inside these rotated squares. To avoid

these complications, the viewing reference frame was rotated by 45 degrees when rendering the colours on the screen, as per Figure 5.1. A triangular lattice can be constructed using square colour samples, as opposed to hexagons or circles. This has the benefit of preserving the shape and size of the individual colour samples in all plane orientations.

Macintosh virtual screen coordinates are integers ranging from -32767 to 32767, with the x-coordinate increasing as one moves to the right and the y-coordinate increasing as one moves downward. These coordinates specify points on a grid. Pixels are enclosed by the lines formed by this grid. Each pixel is associated with the coordinates of the grid point on its upper-left corner. Screen coordinates, depending on the context, are local to a window or global.

The projection from the viewing coordinates (x, y, z) to the screen coordinates is proportional to the coordinates (x', y'), where

$$x' = x - y$$

$$y' = x + y$$

for the square planes and

$$x' = x - y$$

$$y' = z - (x + y)/2$$

for the triangular planes, where z' points out from the screen.

The interaction paradigm for the OSA space is implemented as described in Chapter 5. The implementation is straightforward given the above design, and follows the Macintosh interface guidelines.

Interface consistency is furthered by using QuickDraw-defined window and controls, as well as using interactions common in the Macintosh interface.

Responsiveness is achieved by having all user actions which visibly affect the screen take effect as soon as possible, using QuickDraw primitives whenever possible to maintain consistency.

Permissiveness is best judged by user responses and expectations, as this is a more subjective measure. A better treatment of this is given in Chapter 5.

Bibliography

[App85] Apple Computer, Inc., editor. *Inside Macintosh Volume I.* Addison-Wesley Publishing Co., 1985.

[App88] Apple Computer, Inc., editor. *Inside Macintosh Volume V.* Addison-Wesley Publishing Co., 1988.

[Ben86] K. Blair Benson, editor. *Television Engineering Handbook.* McGraw-Hill, Inc., 1986.

[Bir88] Faber Birren. *Light, Color & Environment.* Schiffer Publishing Ltd., 1988.

[Bra89] D. H. Brainard. Calibration of a computer controlled color monitor. *Color Researach and Application*, 14:23–34, 1989.

[Bri88] Michael H. Brill. Color constancy and color rendering: Concomitant engineering of illuminants and reflectances. *Color Research and Application*, 13:174–179, 1988.

[BW86] D. H. Brainard and B. A. Wandell. Analysis of the retinex theory of color vision. *Journal of the Optical Society of America A*, 3:1651–1661, 1986.

[Com63] Committee on Colorimetry, Optical Society of America. *The Science of Color.* Optical Society of America, 1963.

[Con80] Conrac Division, Conrac Corporation, editor. *Raster Graphics Handbook*. Conrac Division, Conrac Corporation, 1980. Chromaticities for the NTSC and Conrac graphics monitor primaries are given in Figure A3-9a, page A3-17. A discussion of gamma correction factors begins on page 8-17.

[Cow89a] William B. Cowan. Adding colour to the workstation environment. *Graphics Interface '89 Proceedings*, pages 78–85, 1989.

[Cow89b] William B. Cowan. *Colorimetric Properties of Video Monitors*. Notes for a short course presented at the Annual Meeting of the Optical Society of America, Orlando, Florida, October 16, 1989, 1989.

[Eva48] Ralph M. Evans. *An Introduction to Color*. John Wiley and Sons, Inc., 1948.

[Heb55] D. O. Hebb. Drives and the CNS (conceptual nervous system). *Psychological Review*, 62:243–254, 1955.

[Hed88] Patricia Heddell. Color harmony: New applications of existing concepts. *Color Research and Application*, 13:55–57, 1988.

[Hen35] R. H. Henneman. A photometric study of the perception of object color. *Archives of Psychology*, 179:5–88, 1935.

[Hoc83] Robert Hockey, editor. *Stress and Fatigue in Human Performance*. John Wiley & Sons, 1983.

[Int86] International Radio Consultative Committee, Plenary Assembly. *Recommendations and Reports of the CCIR*, volume 11. International Telecommunication Union, 1986. Part 1.

[Jud40] D. B. Judd. Hue, saturation, and lightness of surface colors with chromatic illumination. *Journal of the Optical Society of America*, 30:2–?, 1940.

[Lan59] Edwin H. Land. Experiments in color vision. *Scientific American*, 200:84–94,96,99, 1959.

[Lan77] Edwin H. Land. The retinex theory of color vision. *Scientific American*, 237:108–120,122–123,126,128, 1977.

[LSP89] Robert F. Ladau, Brent K. Smith, and Jennifer Place. *Color in Interior Design and Architecture*. Van Nostrand Reinhold, 1989.

[Mac74] David L. MacAdam. Uniform color scales. *Journal of the Optical Society of America*, 64:1691–1702, 1974.

[Mac78] David L. MacAdam. Colorimetric data for samples of OSA uniform color scales. *Journal of the Optical Society of America*, 68:121–130, 1978.

[Mac90] David L. MacAdam. Redetermination of colors for uniform scales. *Journal of the Optical Society of America A*, 7:113–115, 1990.

[MG80] Gary W. Meyer and Donald F. Greenberg. Perceptual color spaces for computer graphics. *Computer Graphics*, 14:254–261, 1980.

[PS76] J. Pokorny and V. C. Smith. Effect of field size on red-green color mixture equations. *Journal of the Optical Society of America*, 66:705–708, 1976.

[Qui89] Stephen Quiller. *Color Choice*. Watson-Guptill Publications, 1989.

[Rob90] A. R. Robertson. Historical development of CIE recommended color difference equations. *Color Research and Application*, 15:167–170, 1990.

[SCB87] Michael W. Schwarz, William B. Cowan, and John C. Beatty. An experimental comparison of RGB, YIQ, LAB, HSV, and opponent color models. *ACM Transactions on Graphics*, 6:123–158, 1987.

[SV86] Thorstein Seim and Arne Valberg. Towards a uniform color space: A better forumla to describe the Munsell and OSA color scales. *Color Research and Application*, 11:11–24, 1986.

[WF71] G. Wyszecki and G. H. Fielder. Color-difference matches. *Journal of the Optical Societal of America*, 61:1501–1513, 1971.

[Won87] Wucius Wong. *Principles of Color Design.* Van Nostrand Reinhold, 1987.

[WS82] G. Wyszecki and W. S. Stiles. *Color Science: Concepts and Methods, Quantitative Data and Formulae.* John Wiley & Sons, 1982.