Applying Color Science to Computer Graphics

By

Kenneth Paul Fishkin

B.S. (University of Wisconsin, Madison) 1982

THESIS

Submitted in partial satisfaction of the requirements for the degree of

MASTER OF SCIENCE

in

Computer Science

in the

GRADUATE DIVISION

of the

UNIVERSITY OF CALIFORNIA, BERKELEY

Approved:

Brian A. Barsky 9 Dec 83

Chairman **Date**

Paul N. Hilfinger 9 Dec 83

Carlo H. Séquin 9 Dec 83

Committee in Charge

Applying Color Science to Computer Graphics

Kenneth Paul Fishkin

Berkeley Computer Graphics Laboratory
Computer Science Division
University of California, Berkeley
Berkeley, CA 94720

ABSTRACT

Computer graphics is largely concerned with the creation and display of images on a display device. Color display devices support images with very high resolution and dynamic range. As the power of the display devices increases, and the color capacities become more sophisticated, attention to the principles of color science becomes increasingly important. These principles can be applied to many aspects of computer graphics to improve the appearance and correctness of displayed images.

This thesis presents a number of new algorithms in computer graphics; algorithms concerned with display or manipulation of color images. New algorithms are presented which optimally approximate the display of colors which the technology cannot recreate, which quickly translate between one color system and another, which simulate the subtractive mixture of filters and dyes, and which simulate the pigmentary mixture of paints.

December 11, 1983

Acknowledgements

The members of the Berkeley Computer Graphics Lab provided immeasurable help. In addition to technical help, they also provided great personal advice, encouragement, support, and friendship. I'd like to thank Brian Barsky, Tony DeRose, Mark Dippe, John Gross, Rick Speer, Steve Upstill, and Jane Wilhelms, who saw me through from start to finish.

I'd also like to explicitly acknowledge the time, effort, and advice of my thesis readers: Brian Barsky, Paul Hilfinger, and Carlo Sequin.

People both in and out of Computer Graphics were kind enough to lend gratuitous advice and help. Steve Biedermann of Tektronix (formerly of the great University of Wisconsin), George Joblove of Joblove/Kay Inc., and Stu Sechrest and Howard Karloff of Berkeley were particularly kind.

Finally, there were many people and places who made my stay an interesting and enjoyable one. In no particular order, thanks to Blondie's Pizza, Two Guys Who Write Software, rogue 6.1, the U.C. theater, the Oakland A's, Computer Nerds for Social Interaction, the stereo in the lab, Humboldt County, and Sadie Upstill.

How 'bout those Badgers?

TABLE OF CONTENTS

Applying Color Science to Computer Graphics

Kenneth Paul Fishkin

Berkeley Computer Graphics Laboratory
Computer Science Division
University of California, Berkeley
Berkeley, CA 94720

1. INTRODUCTION

"there are no lines in nature - only colors and areas of color. Drawing may have come first in the history of art, but color comes first in human consciousness" - Manet

The field of color computer graphics is by definition intimately bound to color. Color is the *medium of communication* of color computer graphics, influencing and describing all images. The improper use of color can harm the appearance of otherwise valid scenes, yet the role of color science has been little studied and little applied to the field until recently.

This thesis deals with applications in these two realms: color science and color computer graphics.

These two fields are symbiotic to a greater degree than is usually realized. Color science uses computer graphics to confirm, display, and evaluate results of its theories. Computer graphics uses color science on a more pervasive and fundamental level: the correctness and ease and speed of creation of computer graphics images and algorithms are all affected by color science.

Accordingly, the applications in this thesis fall into two categories: using color computer graphics to model color phenomena and applying color science to create and improve algorithms in color computer graphics

The first two chapters introduce these two fields. This thesis is aimed towards a general computer science audience. Therefore, the introduction to color computer graphics focuses only on principles and notation which are unfamiliar to the computer science community at large. Conversely, the color introduction goes into far more detail than would normally be presented in a physics thesis.

After the two introductory chapters, chapter 4 deals with a specific color system translation, a translation used whenever a color is displayed on the display device.

Chapter 5 presents a new algorithm for another color system translation, an algorithm nearly 50% faster than the standard algorithm.

Chapter 6 discusses approximation of colors that the display device cannot recreate. A new algorithm is presented that quickly gives an optimal approximation.

Chapter 7 presents two heuristic algorithms that create spectral and reflectance curves for a color, given only its visual appearance on the display device.

Chapters 8-10 deal with color mixture and implementation of color mixture algorithms. The classical results in additive, subtractive, and pigmentary mixture are reviewed, with the presentation and evaluation of three heuristic algorithms, two simulating subtractive mixture, and one simulating pigmentary mixture.

Chapter 11 concludes the thesis with a brief summary of the new algorithms, and a discussion of open problems.

A small amount of mathematical sophistication is assumed; linear algebra, probability theory, computational geometry, NP-completeness, and abstract algebra are used. Brief explanations are given and references provided for the more esoteric parts.

2. SOME TERMINOLOGY AND NOTATION

This thesis is largely concerned with the display of images on a ***raster device***. These devices are composed of a tightly packed two-dimensional array of individual ***pixels***, each possessing its own color. A pixel is comprised of three ***phosphors*** that are excited by the raster devices' ***electron guns***, and a triple of integers associated with each pixel determines the intensity of that excitation. Each component of the triple is an integer ranging from 0 to N, where N is a function of the particular frame buffer. At Berkeley, an Adage/Ikonas RD3000 frame buffer is used with N = 255.

The frame buffer's pixels will be treated as a two-dimensional array, ***pixel(i,j)***. In accordance with normal array representation, the (0,0) pixel corresponds to the upper left of the display device. The set of pixels on a certain row is termed a ***scanline***.

The two-dimensional pixel array has two operations defined on it:

pixel(i,j) := C ,

sets the pixel at row i, column j, to some color C. Chapter (6) discusses a prefiltering process to ensure that each component of C is between 0 and N.

Similarly,

C := pixel(i,j)

determines the pixel value stored in the frame buffer, and places it in the variable C, assumed to be of type "color".

The process that specifies the set of pixels most closely matching a shape is *scan conversion*. Frame buffers are often used to display geometric shapes, often in polygonal form. The process by which a geometric description of a polygon is *rendered* on a frame buffer is *polygon scan conversion*. When polygons are rendered, their edges will often cover pixels only partially:

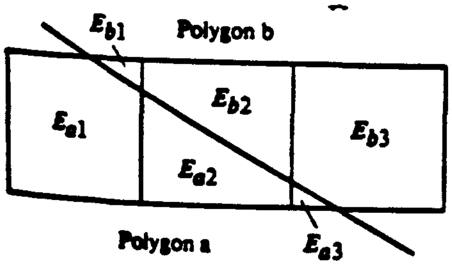

figure 2.1 : a polygon edge partially covers pixels

from Newman and Sproull, pp. 403

If the polygon edge is not properly *sampled*, a defect termed *aliasing* is introduced. *Anti-aliasing* removes spatial high frequency components that cannot be properly rendered with the sampling process inherent in a display with discrete pixels.

In some applications, the user operates directly on the frame buffer representation of the image, as if "painting" on the image directly. Programs of this type are termed *paint programs*.

3. INTRODUCTION TO COLOR SCIENCE

3.1. Introduction

There is one fundamental, simple question at the heart of any discussion of color; the answer and its ramifications comprise this chapter.

"*How can a color be quantitatively measured?*"

This question is by no means as trivial as it seems, for color is not a tangible, objective, physical entity. Rather, it is a *sensation* received by the brain. Indeed, it might seem that measuring color is essentially impossible; pain is a sensation, yet it certainly can't be measured quantitatively!

A restatement of the issue, which also points to an answer, is provided by Heaton:[28]

> "The production of colour is dependent upon light - in the absence of light nothing can be coloured. This sounds a platitude, but it is as well to bear definitely in mind that colour is entirely an abstraction, having no tangible existence" (pp. 1)

The statement "in the absence of light nothing can be coloured" may seem strange; does a red apple in the refrigerator suddenly lose its color when the door is closed? While the apple is the same physical object, it is no longer bathed by the communicating medium of light, and can communicate no sensation of color to the observer. In order for color to be measured, both *an observer* and *a light* must be present.

This intimate connection between color, lights, and observers will permeate this thesis, and especially this chapter. To answer the original question, it seems reasonable, therefore, to attempt to measure color somehow by presenting colored lights to an observer.

3.2. Color matching

One way colors can be specified numerically is by *color matching*. When two colored lights are shone upon a surface, a color is sensed. By varying the strengths of the two lights, a variety of different colors are seen (sensed) by the observer. This set of sensations, then, can be described numerically by a pair of numbers, the relative settings on the two lights that recreated it. For example, the color seen when both lights are at three-fourths strength would be described as (1/2, 1/2), numerically.

This provides a mathematical description of *equivalence of sensations*: the color seen when each light is at three-quarters strength provides the same sensation as that of the original colored light being matched. Similarly, the color seen when each light is at one-quarter strength would differ only in luminance, and would be considered equivalent in "colorfulness" or *chromaticity*.

The two colored lights span, algebraically, a *line segment* of sensed colors. By adding one more light, a *triangle* of colors can be obtained. In general, when n are available, a convex n-gon of colors is obtained. This is one of *Grassman's Laws* of color mixture, discussed in section (3.4).

3.2.1. Negative light

It is well known, for example, that red light shone with blue light produces a magenta (purplish) light. If the primaries of the light mixture system were red and blue, then, magenta would have coordinates of (1/2,1/2).

But what if red *and magenta* are the two primaries? How is blue described? This is done by the obvious but non-intuitive use of *negative* color coordinates. Blue has color coordinates of (-1/2,1/2) with red-magenta primaries. This does not, of course, mean that -1/2 unit of red light is shone on top of magenta light in order to recreate blue light. Rather it means that -(-1/2) = 1/2 unit of red light, when shone with the *blue* light, produces the magenta light.

Equality of color will be used in the mathematical sense hereout, based on an equivalence of *sensation*, not of any objective quality. Two surfaces may differ in every physical property, and yet still produce the same color sensation; this is the phenomenon termed *metamerism*.

3.3. The dimensionality

A tentative step has just been taken into the numerical description of color. If a color can be described by some set of numbers, the next question regards the dimensionality of that set: "Assuming color can be measured numerically, how many dimensions does it span?"

To answer this question, it is nĕcessary to understand something about the workings of the human eye.

3.3.1. Human vision

The human visual system contains three color systems. Each system possesses an *absorption spectrum*, describing the sensitivity of that system to light of each wavelength.

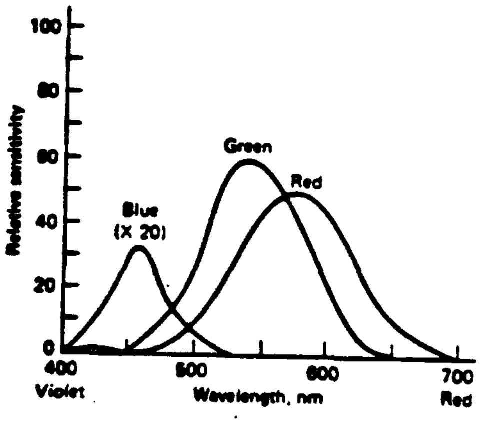

figure 3.1: absorption of the three human color systems

from Foley and Van Dam,[27] *pp. 605*

Each color is sensed by the brain (before any processing) via the transmission of three pieces of information, the amount of stimulation undergone by each color system. The amount of stimulation is a function of two parameters; the spectral composition of the light, and the intensity.[15] For example, the Blue curve above has an absorption percentage of 14% for 520 nm, and 6% for 600 nm. The Blue system transmits the same information upon sensing 1 unit of 520 nm light as it does when sensing (14/6) units of 600 nm light.

The three systems immediately imply the three-dimensionality of color. This fact will be used often in the thesis; colors being represented in any number of three-dimensional terms (points in 3-space, vectors in 3-space, points in 2-space with a weight, elements of a 3-dimensional abelian group, etc.).

At this point, it might appear that the problem of color reproduction can be very easily solved. If the brain receives three pieces of information (one from each system), one need only find three lights, each of which excite exactly one system. Then any color could be easily matched by adjusting the three primary lights to match each of the three levels desired, recreating the sensation of the desired color.

This cannot be done, as demonstrated by the previous picture. There is no wavelength which excites only the green system.

If the responses of the three system at each wavelength are summed, a map of the energy response of the eye to light at each wavelength is developed.

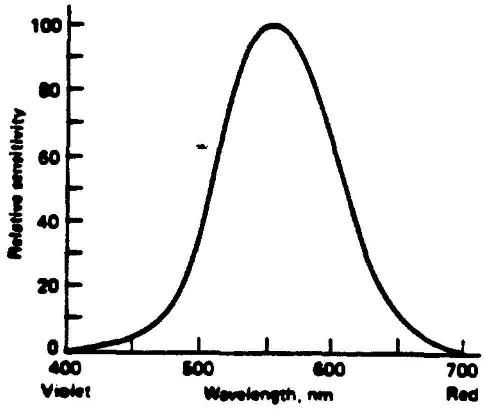

figure 3.2: the curves of figure 3.1, summed

from Foley and Van Dam,[27] pp. 606

The greater the height of the curve, the more receptive the eye to light of that wavelength. Therefore, this curve provides a graph of the perceived *luminance*, or brightness, of each wavelength, a useful curve which will be re-introduced in section (3.10).

3.4. Grassman's Laws

By performing a number of experiments with mixtures of three lights, the first principles of colored light were summarized by Grassman in 1853, and are called *Grassman's Laws*. The notation $A \leftrightarrow B$ means "colored light(s) A produce the same sensation as colored light(s) B". The notation $A + B$ means "colored light A shone upon colored light B".

1) Any colored light A can be expressed as a unique combination of any three other lights:

$$\forall A, B, C, D \ \exists! \ b, c, d \mid A \leftrightarrow bB + cC + dD \tag{3.1}$$

2) A match remains a match, when all lights involved are increased in brightness by the same factor:

$$(A \leftrightarrow B) \Leftrightarrow (kA \leftrightarrow kB), \ \forall k \geq 0 \tag{3.2}$$

3) If a light A is matched by some other combination of colored lights, that combination can be used in place of A in all cases, with identical effect. This law forms the basis for modern color reproduction.

Grassman stated one assumption, as well:

1) The luminance produced by the additive mixture of a number of lights is the sum of the luminances provided separately by each of the lights. This assumption forms the basis of modern photometry.

These laws can be shown to mathematically categorize color as a three-dimensional vector space: three-dimensional due to the nature of the human eye, a vector space due to Grassman's second and third laws.

Any color space whose primaries comprise three basis vectors in this three-dimensional space are termed *additive* color spaces.

From the three laws, Grassman proved that the color formed by the mixture of colored lights is the weighted average of the colors of the constituent lights. The weights correspond to the relative proportions of the light in the mixture.

It has been shown that under certain unusual conditions Grassman's Laws may not hold, but they hold over a sufficiently large range of conditions with sufficiently large precision that their validity is generally accepted. [†]

3.5. Chromaticity

The color triangle described in section (3.2) is, obviously, two-dimensional. As demonstrated in the previous section, color (to humans) is a three-dimensional sensation. Therefore, it follows that one dimension is being left out. This dimension is the dimension of ***brightness*** or ***luminance***, since only the ***relative***, and not the ***absolute*** amounts of the primaries are specified in the triangle.

Since these triangles describe the *colorfulness* or *chromaticity* of the colored light, they are called *chromaticity diagrams*, and the coordinates of a color in such a map its *chromaticity coordinates*.

3.5.1. The CIE chromaticity diagram

By using exactly the procedure described in section (3.2), colorists derived the famous chromaticity diagram of color visible to the human eye seen below. This diagram is termed the *CIE chromaticity diagram* for the *CIE* ("Commision Internationale de l'Eclairage") international society of colorists.

[†] the effect of the surrounding background colors on the color perceived for the foreground color is also ignored

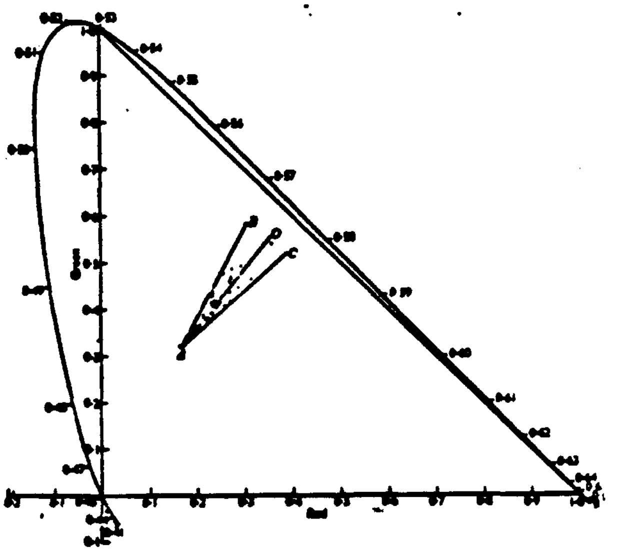

figure 3.3: the CIE chromaticity diagram

from Wright,[63] pp. 95

The set of colors found on the boundary is termed the *spectral locus*, consisting of all colors that contain only a single spectral wavelength.

This figure is perhaps the most important in the thesis. It maps any color pèrceivable by the human eye (projected onto the luminance plane) into a two-dimensional point. There is one feature of great importance in this figure: *it cannot be inscribed by a triangle of visible colors.*

While the set of visible colors can be inscribed in a triangle, this triangle *cannot* contain, as vertices, colors that are themselves realizable. There will always be colors which are not obtainable as a mixture of any given three primaries. Similarly, a triangle can be inscribed in the chromaticity diagram, but this triangle will not contain all the realizable colors.

3.6. rgb space

This leads naturally to the next question: "if *all* of color space can't be spanned by any three primaries, which choice of primaries would *maximize* the percentage of color space spanned?". In fact, those three primaries are exactly red, green, and blue!

The intuitive importance of these three colors is thus confirmed mathematically. The precise placement of red, green, and blue actually varies slightly from this theoretical solution, for technical reasons dealing with the luminance of available lights and the sensitivity of the eye to different areas of the spectrum. The exact wavelengths chosen are shown below:

Wavelengths of C.I.E. spectral red, green, and blue (nm)	
red	700.0
green	546.1
blue	435.8

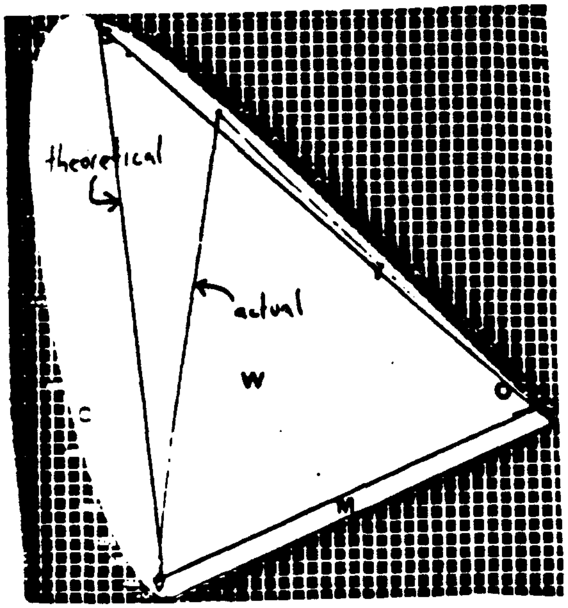

figure 3.4: the theoretical and actual placement of CIE red, green, and blue

The triangle of colors spanned by C.I.E. red (r), green (g), and blue (b) is termed ***rgb space***. The ***rgb*** coordinates describe the ***relative*** intensities of the three lights.

In *RGB* space, the ***absolute*** values are specified. As the absolute values represent the "absolute stimulus", and there are three of them, the absolute values are termed the ***tristimulus coordinates*** of a color.

The RGB value of a color, then, represents the ***absolute stimulus*** amounts of Red, Green, and Blue light needed to match the sensation of some given color. Red, Green, and Blue are defined as monochromatic light of 700, 546.1, and 435.8 nanometers, respectively. Since often only the ***relative*** amounts of the light are desired, the luminance can be factored out, resulting in ***rgb space***:

$$\begin{aligned} r &= R/(R+G+B) \ , \\ g &= G/(R+G+B) \ , \\ b &= B/(R+G+B) = 1-r-g \end{aligned} \tag{3.4}$$

This convention is used for other color spaces as well; lower case refers to chromaticity (relative) coordinates, while upper case refers to tristimulus (absolute) coordinates.

By the previous figure, the set of colors spanned by rgb space is

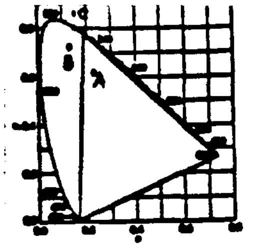

figure 3.5: CIE space and rgb space

from Evans, pp. 238

point A can be matched by a combination of red, green, and blue

point B is visible to the eye, but cannot be matched by red, green, and blue

point C is not visible to the human eye.

3.7. Unwinding the spectral locus

The preceding figure defines the placement of every point in the spectral locus. However, it does so geometrically, not algebraically; it is not easy to determine the chromaticity coordinates of a color of wavelength λ.

This can be done by "unwinding" the spectral locus. Given a triad of Red, Green, and Blue lights of some fixed luminances, we can take every spectral color of unit luminance and obtain a triple of numbers by the process of section (3.2). By repeating this process across all wavelengths, three curves are obtained; one for each of Red, Green, and Blue. These curves are termed the *color-matching curves* for a given set of primaries, since they specify how much of each colored primary is needed to match any given spectral color. In the case of *rgb* primaries, some of the curves dip below zero.

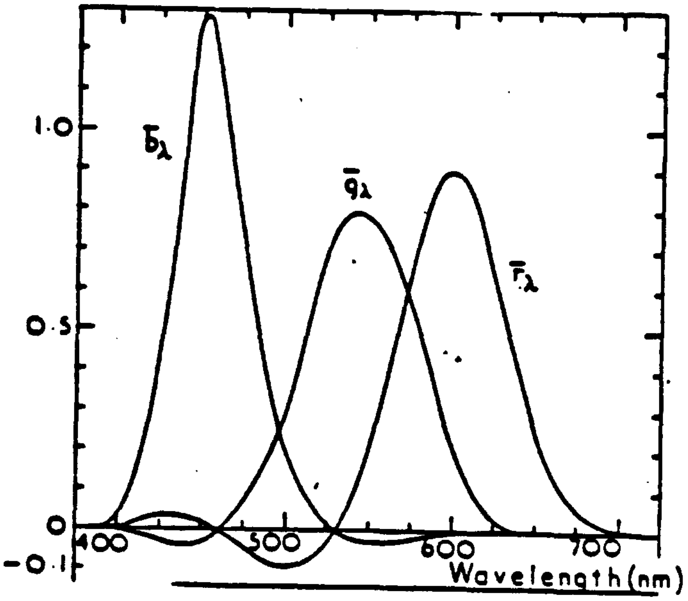

figure 3.6: the rgb color-matching curves

from Hunt,[31] pp. 77

3.8. Finding the RGB value of a color

The next question is one of physical measurement: given a colored surface or light, how is its RGB value computed?

This could be done by the procedure mentioned in section (3.2): A triad of colored lights could be obtained, manually adjusted until they recreated the appearance of the desired color, and the readings of the dials then read. In practice, this is time-consuming, inefficient, and inaccurate.

Usually, a *spectrophotometer* is used to obtain a reflectance at all visible wavelengths (see figure (3.7)).

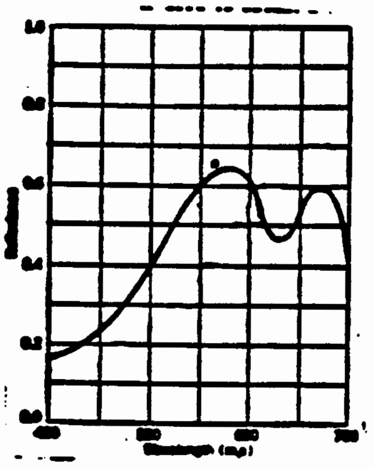

figure 3.7: a sample reflectance curve

from Evans, pp. 95

3.8.1. Reflectance vs. transmittance

This sample reflectance curve introduces an important issue, the dual nature of color science and color measurement.

The curve in figure (3.7) has an upper bound of 1. This represents the fact that a surface can never reflect more light than is shone upon it.

If one were to hold a flashlight one foot from the surface of the sun, however, the sun's surface would nevertheless be brighter than the light of the flashlight! This is because the sun *emits* light, being a *self-luminous* body.

Accordingly, there are *two* curves used in color science. *Reflectance* curves, as mentioned above, represent the percentage of light which is *reflected* at each wavelength from a surface.

Transmittance curves, conversely, represent the light energies which are *transmitted* by the light source at each wavelength. There is no upper bound on these curves; within the realm of physical possibility, the sun's energy can become infinitely great.

In addition, in optics, *internal transmittance* curves are sometimes used to describe the *percentage* of light which is transmitted *through* a filter.

3.8.2. Translation from curve to RGB

The translation from spectral reflectance (or transmittance) curve to RGB value is non-trivial, and can be visualized concretely by considering the simple example of monochromatic light.

Consider the example of a spectral color, caused by light of wavelength 550 nm. Then the transmittance curve is

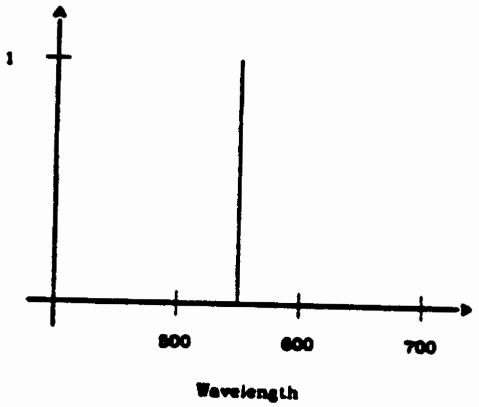

figure 3.8: monochromatic transmittance

Multiplying the rgb color matching curves (figure (3.6)) by the preceding curve, figure (3.9) is obtained:

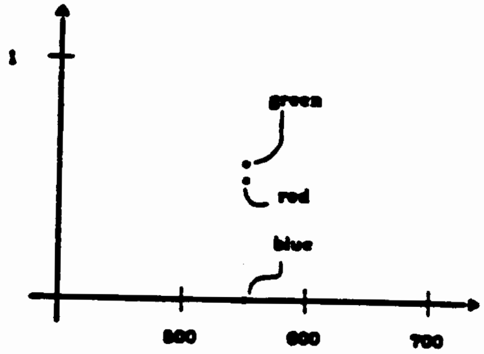

*figure 3.9: transmittance *matching functions*

The heights of the three multiplied curves exactly match the desired RGB value.

In general, the RGB values for any transmitted light are computed as[32]

$$R = \frac{\int_{\lambda=400}^{700} S_\lambda \bar{r}_\lambda \, d\lambda}{\int_{\lambda=400}^{700} \bar{r}_\lambda \, d\lambda} \qquad (3.3)$$

$$G = \frac{\int_{\lambda=400}^{700} S_\lambda \bar{g}_\lambda \, d\lambda}{\int_{\lambda=400}^{700} \bar{g}_\lambda \, d\lambda}$$

$$B = \frac{\int_{\lambda=400}^{700} S_\lambda \bar{b}_\lambda \, d\lambda}{\int_{\lambda=400}^{700} \bar{b}_\lambda \, d\lambda}$$

where

S_λ is the spectral energy of the stimulus at wavelength λ.

$\bar{r}_\lambda$, $\bar{g}_\lambda$, $\bar{b}_\lambda$ are the red, green, and blue color matching functions.

In case of a different color system, with differing matching functions, only the new matching functions need be substituted into the above equations.

$\bar{r}_\lambda$, $\bar{g}_\lambda$, $\bar{b}_\lambda$ have negative values in places.

3.8.3. RGB value of a surface

Only half of the problem has been solved; computing RGB from spectral *transmittance*. The case of spectral *reflectance* is handled below.

Consider the red apple mentioned at the start of this chapter. Its color can be measured by shining white light onto it, and then measuring the light that is reflected off the apple. In that case, the procedure of the preceding section could be used.

But now, suppose the light shining on the apple is increased in intensity. Naturally, the apple now appears brighter. However, the surface of the apple has not changed its properties. Therefore, the tri-stimulus values of surface colors are ***divided by the luminance of the illuminating light.*** This result will be restated mathematically in section (3.10), when a more precise definition of luminance is given.

3.8.4. Summary

A triple of numbers representing a color has two markedly different meanings, depending on whether it represents transmitted (*emitted color*) or reflected color (*object color*).

This difference is often glossed over, due to their similarities. For example, a green-colored surface can be obtained either by shining a green light source on a white surface, or a white light source on a green surface.

3.9. Advantages and disadvantages of RGB space

RGB space has several advantages: RGB value can be computed from spectral curves, RGB space forms a three-dimensional vector space, and physical recreation of color from RGB value is simple.

3.9.1. Disadvantages of RGB space

RGB space has two colorimetric disadvantages. As mentioned in section (3.7), the color-matching curves are not non-negative, which could make computation of RGB values from spectral curves awkward. In addition, the three color-matching curves are asymmetric in area, a skewing factor that must be corrected for (see equation (3.3)).

From the standpoint of user interaction, RGB space is also unfortunate. The *perceptual distance* between two colors is a relative measure of the amount by which two colors appear to differ to a human observer. These distances can be determined via psychometric just-noticeable-difference (*jnd*) techniques. Equal distances in RGB space do not correspond to equal perceptual difference. For example, the RGB distance between magenta (1,0,1) and blue (0,0,1) equals that between magenta (1,0,1) and white (1,1,1). Yet magenta looks quite similar to blue, while magenta and white are strikingly dissimilar.

Similarly, it is difficult (especially for a naive user) to conceive of a color in terms of its RGB value. What are the RGB values of grape? of tan? of banana yellow? All but the simplest colors are difficult to specify in the RGB system. This may be one of the reasons why computer graphics images that are highly complex in every other aspect†, tend to have very primitive and simple colors.

3.9.2. RGB space versus monitor space

Despite overwhelming assumption to the contrary by graphics programmers, RGB space is *NOT* the color space used by present monitors![63, 31]

Monitors contain three *electron guns*. There is an array of phosphors on the screen, arranged in groups of three. Each phosphor in the group is excited by a different gun. By controlling the intensity of the beam from the electron gun, the amount of light emitted by the phosphor is controlled.

If the red, green, and blue phosphors had exactly the colors of C.I.E. Red, Green, and Blue, this would in fact be identical to the RGB system. As the figure below demonstrates, this is not the case. The set of colors achievable by the TV

† for example, the "refracting sphere" sequence by Whitted, and the "Growth" sequence by Kawaguchi

monitor is a ***proper subset*** of RGB space.

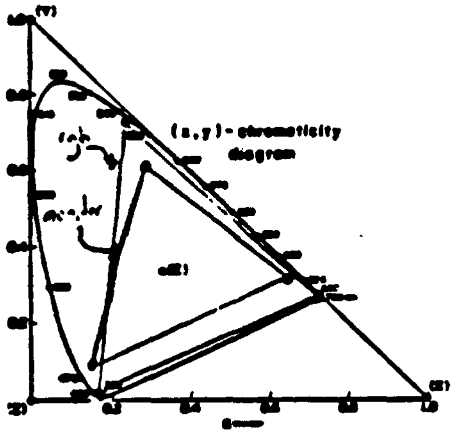

figure 3.10: CIE vs RGB vs monitor space

The color space spanned by the monitor, then, is simply another additive color space, with a different set of primaries from the RGB primaries. The monitor color space will be referred to as ***mRGB space***, and the value of gun excitation which produces a color as its ***mRGB value***. The ***mrgb*** space of chromaticity coordinates is defined in the canonical way, obtained by dividing the mRGB components (***mR, mG, and mB***) by their sum.

As can be seen from the preceding figure, the RGB coordinates of a color are somewhat similar to the excitation levels of red, green, and blue electron guns on a raster display device needed to reproduce that color, but this they are not identical.

The mRGB system, just as the RGB, deals with transmitted light, the amount of light transmitted by each of the three primaries. Unlike RGB space for transmitted light, however, mRGB space is a bounded space. This is a constraint imposed by the physical device, the guns possessing a maximum energy. The problem of displaying a color possessing a luminance too great for monitor display is discussed in chapter (6).

3.10. The XYZ System

The disadvantages of the RGB system for colorimetry are substantial; the negative (and unbalanced) matching functions, and the lack of an easy luminance metric make it impractical for colorimetric use.

In 1931, the CIE proposed a color system to mitigate many of the failings of the RGB system. This system uses three different primaries, the *X*, *Y*, and *Z* axes to span color space; hence the term *XYZ space*.

There are, of course, an infinite number of primaries that could have been chosen. The XYZ primaries were chosen very intelligently to satisfy a number of criteria, including:

1) The primaries should span *all* of color space in non-negative combinations. This immediately implies that the primaries are *not visible colors* (see figure (3.5)). This concept is a simple one, mathematically; any set of three (linearly independent) basis vectors will span color space; there is no reason why the color points corresponding to those vectors need be inside color space.

Again, in the RGB system the R primary, for example, represents an amount of physical red light. In the XYZ system, however, no single primary represents a physical light; the primaries simply *span* the space of visible color in non-negative combinations.

This non-negativity immediately implies non-negative color-matching functions, which greatly aids the reconstruction of XYZ values from spectral curves.

2) The total area under each of the color-matching functions should be equal. Otherwise, a normalizing division must be done when computing XYZ values from spectral curves, as in equation (3.3).

3) Color essentially consists of two different qualities: chromaticity (two-dimensional) and luminance (one-dimensional). Tying one of the three basis vectors directly to luminance is desirable.

The CIE satisfied this requirement in an ingenious way. In the discussion of color vision, a function was derived encoding perceived luminance as a function of wavelength (figure 3.2). If one of the color-matching functions is *made equal* to this curve, it will, by definition, capture the luminance of color.

The CIE Y axis was chosen in precisely this manner, and represents the luminance of color. The X and Z axes were chosen in order to fit other requirements in conjunction with the existing choice of Y, and possess *no* luminance.[38, 63] The three axes so chosen specify the well-known XYZ color-matching functions:

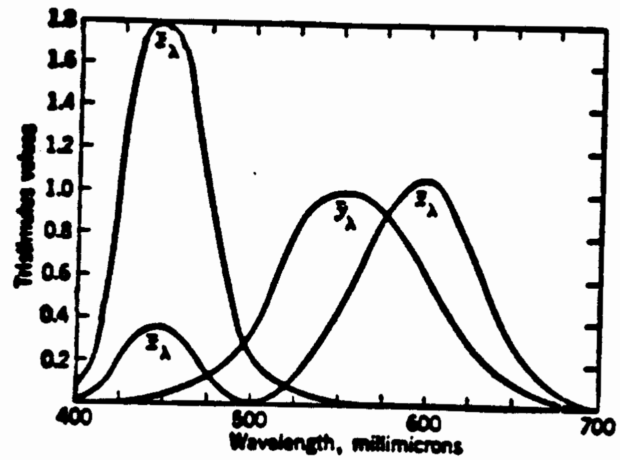

figure 3.11: the XYZ color-matching functions

from Judd,[39] pp. 43

For a further discussion of the foundations of the XYZ system, see Hunt,[31] Judd,[39] Hunter,[32] Wyszecki,[64] or especially Wright.[63]

The XYZ values are computed from spectral transmittance by (from Hunter)[32]

$$X = \int_{\lambda=400}^{700} S_\lambda \, \bar{y}_\lambda \; d\lambda \tag{3.5}$$

$$Y = \int_{\lambda=400}^{700} S_\lambda \, \bar{y}_\lambda \; d\lambda$$

$$Z = \int_{\lambda=400}^{700} S_\lambda \, \bar{y}_\lambda \; d\lambda$$

where

S_λ is the spectral energy of the stimulus at wavelength λ, and $\bar{x}_\lambda$, $\bar{y}_\lambda$, and $\bar{z}_\lambda$ represent the color matching functions.

The XYZ values are computed similarly from spectral reflectance by[32]

$$X = \frac{\int_\lambda E_\lambda \; R_\lambda \; \bar{x}_\lambda \; d\lambda}{\int_\lambda E_\lambda \; \bar{y}_\lambda \; d\lambda} \tag{3.6}$$

$$Y = \frac{\int_\lambda E_\lambda \; R_\lambda \; \bar{y}_\lambda \; d\lambda}{\int_\lambda E_\lambda \; \bar{y}_\lambda \; d\lambda}$$

$$Z = \frac{\int_\lambda E_\lambda \; R_\lambda \; \bar{z}_\lambda \; d\lambda}{\int_\lambda E_\lambda \; \bar{z}_\lambda \; d\lambda}$$

where

E_λ is the energy of the illuminant at wavelength λ

R_λ is the percent reflectance of the color at that wavelength λ.

These equations demonstrate the division by scene luminance described in section (3.8.3). As the intensity of the illuminant is increased so do the denominators, resulting in an equivalent XYZ value.

By definition, the Y for the perfect white is thus 1.0.

3.10.1. xyY space

Just as in RGB space, the tristimulus XYZ values can be mapped to their xyz chromaticity coordinates by

$$x = X/(X+Y+Z) \;, \tag{3.7}$$

$$y = Y/(X+Y+Z) \;,$$

$$z = Z/(X+Y+Z) \; = 1-x-y$$

Any two of the three chromaticity coordinates, in conjunction with the luminance Y, thus describe any color, while (*xyz*) does not. This *xyY space* has the distinct advantage that two dimensions (x and y) uniquely describe the chromaticity, while the third (Y) describes the luminance.

3.10.2. Psychophysical systems

None of the color spaces mentioned so far have necessarily possessed any relationship between Euclidean distance in the color space, and perceptual distance.

In 1937, MacAdam proposed a *uvY* color system, obtained as an *algebraic* transformation from xyY space; the transformation is explicitly designed such that equal distances in uv space correspond to equal distances in perceptual space. This system did not extend this property into the luminance dimension, and 1963 the C.I.E. proposed the *U*V*W** color system, which maintains this property in all three dimensions (the L*a*b* is another modification of this system). The defining transformation required to transform U*V*W* coordinates to or from RGB coordinates is computationally expensive, and the system has been little used to the author's knowledge. Two similar color systems, the CIELUV and CIELAB color systems, also attempt to map color to a color space with Euclidean perceptual distance.

There is some uncertainty regarding the semantic validity of these color spaces. Color consists of two qualities, chromaticity and luminance. While it appears reasonable to create mappings for chromaticity distance and luminance, it may be meaningless to *connect* these mappings into the third dimension, a point made excellently by Evans:[23]

> "At first sight [such a color space] would appear to be logical and worth while and seems

> not only possible but plain common sense. There are many reasons to believe, however, that such an arrangement is a purely imaginary concept which could be realized in a practical way only by so restricting its meaning and application as to make it quite useless. In the first place a just perceptible difference in one part of the color solid may not have the same meaning that it does in another. Almost certainly four just perceptible differences in brightness are a different *kind* of difference than four in hue, etc. In other words, whereas such an arrangement has a very logical-sounding basis, it does not necessarily lead to useful concepts. What is really wanted is an arrangement of colors having the property that a distance in one region indicates a color difference which looks like the same difference as that indicated by a line of the same length somewhere else. Much evidence seems to be accumulating that such an arrangement is generally an impossibility" - pp. 224

3.11. The YIQ system

Transmission of color television signals required another choice of color system. In the United States, the NTSC (National Television Systems Committee) chose a set of basis vectors such that one represented the luminance of the color (Y), and another lay along the axis of flesh tones (I). The third axis (Q), was simply chosen to be have a vector component sum of 0, and to be independent of the other two.

This *YIQ* system is a linear transform from RGB space, Y encoding the luminance information. The advantage of this decision is that the same signal can be sent to both color and black-and-white displays. Black-and-white sets use the Y signal to control gray level, and ignore I and Q.

The Y axes in the YIQ and XYZ system are identical.[27] The "YIQ Y" is defined as[27]

$$Y_{YIQ} = .30mR + .59mG + .11mB$$

for a standard color monitor, while the "XYZ Y" is defined as

$$Y_{XYZ} = .17R + .81G + .01B$$

for CIE Red, Green, and Blue. The axes *are* identical; the different coefficients describe the location of the Y axis as combinations of two different sets of basis vectors; { mR, mG, mB} for YIQ Y, { R, G, B } for XYZ Y.

3.11.1. Monitor correction

When a YIQ signal is received by an NTSC color television, it is transformed via a matrix transform into a mRGB value. These mRGB values cannot be passed directly as gun voltages, however, since the light output by the tube is proportional to a power γ of the applied voltage.

The process that aligns linear increases of mR,mG, or mB with linear increases in light output is accordingly called *gamma correction*, and is typically applied by taking the γ'th root of the mRGB values. γ commonly ranges between 2.5 and 3.0, depending on the monitor.[49]

3.12. Dominant Wavelength and Purity

Any color consists of varying amounts of light across wavelengths - these specify the color precisely, as in the XYZ system. One way to specify a color's chromaticity is by its *dominant wavelength* (Λ), and *purity* (ρ)

Λ is the particular wavelength of light the color emits to the greatest degree. Given Λ, a color can be recreated by a mixture of white light with pure light of wavelength Λ. ρ is the percentage to which that wavelength predominates in the curve.[32]

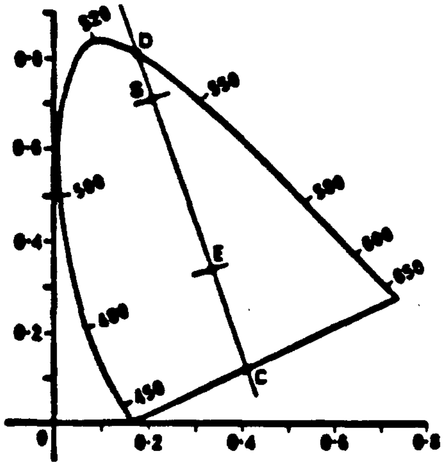

figure 3.12: the Λρ system

from Osborne,[51] *pp. 93*

This Λρ system is very naturally described geometrically in an xyY chromaticity diagram. If a line is drawn containing the white point E and the chromaticity coordinates of the desired color S, it will intersect the spectral locus at some pure color D of wavelength Λ The purity is defined geometrically as

$$p = \frac{ES}{ED}$$

The purity is therefore the ratio of two distances: the distance from the white point to the given color, as opposed to the distance from the white point to the spectrum locus.[32]

Λ corresponds exactly to hue, and ρ to saturation. This system thus provides a bridge between the purely technical xyY and XYZ color spaces, and the more intuitive hue and saturation-based color spaces (see below).

3.13. Perceptual color spaces

The color spaces described have been rigid mathematical spaces, used for scientific and engineering purposes. Artists and painters have also used color for hundreds of years, and have developed color systems of their own. These color systems are often termed *perceptual* color spaces since they are considered to align more closely with intuitive human conception of how colors are ordered.

Since their foundation is intuitive, not physical, the arrangement and placement of colors within these spaces has a good deal of liberality.

There are two main artistic perceptual color spaces: the *HSV* (Hue, Saturation, and Value) and *HSL* (Hue, Saturation, and Lightness) spaces.

The *hue* of a color is loosely defined as "the main quality factor in color...the essential element that leads us to name it red or green".[23] Intuitively, it corresponds to the basic "color orientation" ("reddish", "bluish", etc.) of the color.

The *saturation* of a color is the percentage of hue in a color,[23] or the "density" of the color. For example, grayish-pink has the same hue as red, but a lower saturation. Colors with zero saturation (grays) are termed *achromatic*, and those with non-zero saturation *chromatic*. The *chromaticity* of a color is usually defined as its hue and its saturation. Similar to the previously defined chromaticity coordinates, these are two color dimensions that describe a color without regard to the light intensity it emits.

Hues are easily and naturally arranged in a circle. Red, green, and blue are equally spaced around the edge of the circle. Other colors are then located around the circle according to their red/green/blue proportions; the ratio of the two largest components each decreased by the smallest component. A color that is, for example, one-third red and two-thirds green is located one-third of the way from red to green. All colors that only contain one or two of the three primaries can thus be located around the circle, according to the proportion of their primaries. This is the basic color arrangement we have all seen in paint stores.

Note that virtually any arrangement of hues in the circle can be created, depending on the original spacing of red, green, and blue. The canonical arrangement has red, green, and blue equispaced around the circle, but any one-to-one and onto map of hue to circular angle is valid, since it preserves the descriptive power of the arrangement.[37] Hunter for example,[32] uses a circle with green twice as close to blue as red, Fishkin[25] one with green three times as close to blue as red, and Joblove[37] performs a sinusoidal mapping to eliminate Mach-banding. The particular intuition of the user and the particular requirements of the application determine the mapping.

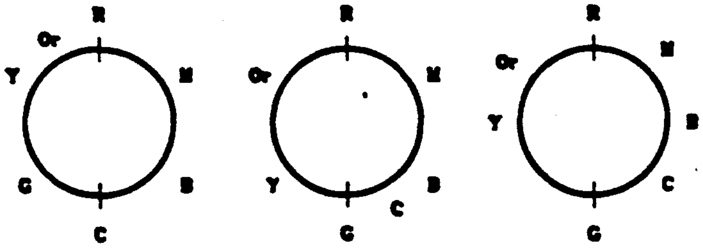

figure 3.13: different hue circles,
from SIGGRAPH Core,[13] Fishkin,[25] and Hunter[32].

The amount of the third (smallest) primary determines the distance of the color from the center of the circle - the distance is defined as proportional to the color's saturation. For example, pure red would be on the edge of the circle, and pink on the same angle, but a little closer to the center.

To reiterate, the polar angle of a color is determined by the ratio of its two largest primaries, and its radius by its saturation. This polar coordinate system specifies the placement in a plane of a color in both the HSL and HSV systems. The difference is in the choice of the third dimension.

3.13.1. The HSV system

In the HSV system, the non-blackness of color, its *value*, is used for the third dimension. The value is defined in RGB terms as max(R,G,B).[57] For example, black has a value of 0, grey a value of .5, and red, green, blue, and white all have values of 1. This color system is represented by a cone, due to the singularity of black, the only color with a value of 0. The conical cross-sections grow gradually smaller as the singularity of black is approached. The radius of a color from the center on a cross-section defines the color's *chroma*. The saturation of a color is usually defined as the ratio of the chroma to the maximum radius at that value.

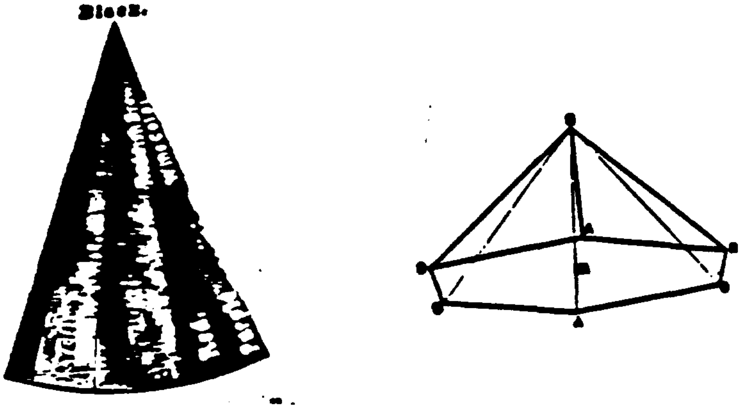

figure 3.14: the HSV systems
of Von Bezold,[5] *pp. 106, and Kuppers,*[42] *pp. 165*

This color system is used often by artists, mainly because the base (value = 1) contains most of the "important" colors, arranged in an intuitive format.

Value is simply defined in RGB terms, as stated above, as max(R,G,B). Imagine the colorcube described before, projected along its main diagonal (the greyscale) onto a plane perpendicular to this diagonal. A hexagonal outline results (think of the silhouette of a isometric drawing of a cube). This is exactly the set of all colors with a value of 1. Similarly, the set of all RGB colors with maximum component less than or equal to some arbitrary value **V**, will form a smaller colorcube. The projection of this set possesses a hexagonal outline, and contains exactly the set of colors with a value of **V**.

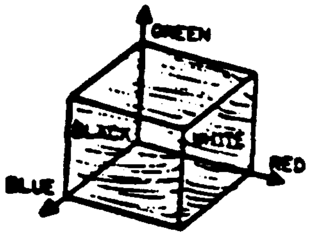

figure 3.15: value = 1 in RGB space

For this reason, in some computer graphics literature[27, 58, 57] the conical cross-sections of perceptual color spaces are often depicted as hexagons, while other authors[5, 23, 32, 36, 63] use the more intuitive circular arrangement described above. While the hexcone may appear to be preferable in light of the geometric argument presented above, the fact that the set of colors with a given value has a hexagonal silhouette in RGB space does not imply that the same set of colors has a hexagonal silhouette in HSV space.

3.13.2. The HSL System

In the HSV system, the third dimension measures the non-blackness of a color. In the HSL color system, the third dimension represents the whiteness or *lightness* of the color. This is defined in RGB terms as (max(R,G,B) + min(R,G,B) / 2).[13] For example, black has a lightness of 0, red, green, and blue all have lightnesses of .5, and white has a lightness of 1.

This color system is a double cone, due to the two singular points of black and white.

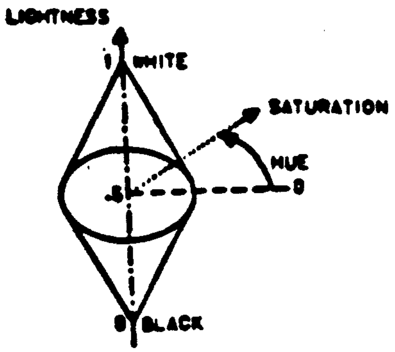

figure 3.16: HSL space

from the '79 Core,[13] pp. III-7

Smith[59] feels HSL space is less intuitive than HSV space, due to the double-cone arrangement, while others[38] prefer HSL. However, HSL does have the feature that grey, not white, lies at the center of the primary plane, which is desirable for some applications (see section (10.4)).[25]

3.14. Conclusions

The computer graphics programmer can choose from many different color spaces. The desires to correct monitor deficiencies, to provide users with an intuitive system, and to simulate physical color phenomena guide this choice. Image appearance, correctness, and description are all affected by this choice. There is no single color system that is best for all applications; as with programming languages, the needs of the application suggest the appropriate tool.

4. Transforming between XYZ and RGB

Since both the XYZ and RGB systems are additive and describe all colors by their position in vector spaces spanned by three basis vectors, the two systems are related by an equation of the form

$$\begin{bmatrix} R \\ G \\ B \end{bmatrix} = \overline{\mathbf{M}} \begin{bmatrix} X \\ Y \\ Z \end{bmatrix}$$

The derivation of this transformation must be done with some care, and is often glossed over or ignored. Accordingly, in this section the transformation will be derived in some detail.

4.0.1. Translating from xyz to rgb

The task of translating from XYZ to RGB can be approached by solving a considerably simpler problem; translating from xyz to rgb. By this translation the ***tristumulus coordinates*** have been replaced by the ***chromaticity coordinates*** by the familiar transform of sections (3.10) and (3.8).

Xyz to rgb translation can be performed very simply, by referring to the chromaticity coordinates of standard red, green, and blue.

chromaticity coordinates of red, green, and blue				
name	λ*	x	y	z
red	700.0	.73467	.26533	0.0
green	546.1	.27376	.71741	.00883
blue	435.8	.16658	.00886	.82456

*λ in nm.

The xyz to rgb transform can be obtained by inverting the transformation

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} .73467 & .27376 & .16658 \\ .26553 & .71741 & .00886 \\ 0.0 & .00883 & .8246 \end{bmatrix} \begin{bmatrix} r \\ g \\ b \end{bmatrix} \qquad (4.1)$$

Yielding

$$\begin{bmatrix} r \\ g \\ b \end{bmatrix} = \begin{bmatrix} 1.5771 & -.5980 & -.3122 \\ -.5834 & 1.615 & .1005 \\ .0062 & -.0173 & 1.2117 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix}$$

This equation is easily mistaken for the final answer. The chromaticity diagrams are *planar slices* of a three dimensional vector space; the transformation given above maps *lines* onto *lines* in the three-dimensional vector space, *not* the desired map of *points* onto points.

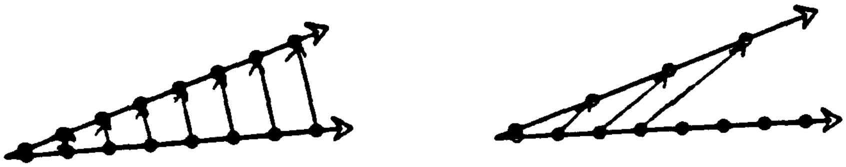

figure 4.1: two different mappings preserving chromaticity

In other words, the xyz to rgb map is a *two*-dimensional one, while the desired XYZ to RGB map will be *three* dimensional. Therefore, the rgb to xyz transforming equation above is under-specified for a RGB to XYZ translation; a three-dimensional map cannot be derived solely from a two-dimensional one. While it guarantees that the chromaticities will be mapped correctly, more information must be used in order to complete the transformation.

4.0.2. Completing the transformation

The chromaticity transformation can be extended to a tristimulus transformation by noting that any linear multiple of the chromaticity mappings forms a correct tristimulus transformation. For example, mapping $(1,0,0)_{RGB}$ to $(.73467,.26533,0)_{XYZ}$ will obviously maintain the correct chromaticity. But by definition of the XYZ to xyz transform, so will $(.73467k,.26533k,0)_{XYZ}$ for any non-zero k.

The transformation can thus be written as

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} \rho \begin{pmatrix} .73467 \\ .26553 \\ 0.0 \end{pmatrix} & \gamma \begin{pmatrix} .27376 \\ .71741 \\ .00883 \end{pmatrix} & \beta \begin{pmatrix} .16658 \\ .00886 \\ .82456 \end{pmatrix} \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix} \tag{4.1.5}$$

This equation is solved in ρ,γ,β by adding two conditions.[63, 10, 31] First, in both systems achromatic light has equal tristimulus coordinates in each component. For all k, $(k,k,k)_{XYZ}$ must map to $(m,m,m)_{RGB}$, for some m.

Second, the color point representing one trichromatic unit of white in the XYZ system is constrained to transform to a color point representing one trichromatic unit of white in the RGB system.[10, 31, 63] The number of trichromatic units is simply the sum of the components. Therefore, in conjunction with the first claim, this claim implies that $(1/3,1/3,1/3)_{RGB}$ maps to $(1/3,1/3,1/3)_{XYZ}$:

$$\begin{bmatrix} \frac{1}{3} \\ \frac{1}{3} \\ \frac{1}{3} \end{bmatrix} = \begin{bmatrix} \rho \begin{pmatrix} .73467 \\ .26553 \\ 0.0 \end{pmatrix} & \gamma \begin{pmatrix} .27376 \\ .71741 \\ .00883 \end{pmatrix} & \beta \begin{pmatrix} .16658 \\ .00886 \\ .82456 \end{pmatrix} \end{bmatrix} \begin{bmatrix} \frac{1}{3} \\ \frac{1}{3} \\ \frac{1}{3} \end{bmatrix} \tag{4.2}$$

This can be trivially re-written as a linear system in three unknowns:

$$\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} .73467 & .27376 & .16658 \\ .26553 & .71741 & .00886 \\ 0.0 & .00883 & .82456 \end{bmatrix} \begin{bmatrix} \rho \\ \gamma \\ \beta \end{bmatrix}$$

Solving this linear system yields $\rho = .66694$, $\gamma = 1.1324$, and $\beta = 1.2006$. Substitution of these values into equation (4.2) yields the final system:

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} .48989 & .31001 & .20000 \\ .176696 & .81240 & .01064 \\ 0 & .01000 & .98999 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix} \tag{4.3}$$

The inverse transformation is then

$$\begin{bmatrix} R \\ G \\ B \end{bmatrix} = \begin{bmatrix} 2.3647 & -.51515 & .00520 \\ -.896656 & .14264 & -.01441 \\ -.46808 & .08874 & 1.00921 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \end{bmatrix}$$

This procedure is a specific case of a more general process. Equation (4.1.5) is a system of three equations in nine unknowns. By supplying the values of six (R,G,B,X,Y,Z) of the nine unknowns, the system was solved. Any set of constraints which supply six degrees of information will also serve.

4.1. Transforming between mRGB and XYZ

Reviewing the process of the previous section, only two sets of data are required in order to translate between RGB and XYZ: the chromaticities of the primaries, and the location of the "white point". Translation between mRGB and XYZ is simply another instantiation of this process.

The chromaticity coordinates of the three phosphors will be denoted as CC_{mr} , CC_{mg} , and CC_{mb}. These coordinates can be obtained by direct measurement from the monitor involved, or, with less accuracy, from the manufacturer (see Cowan[16] or Meyer[49] for details).

The luminance of the "white point" will be referred to as L_w. This can only be reliably derived from spectrophotometric measurement of the monitor involved.

4.1.1. Translation between mrgb and xyz

Knowledge of the chromaticity coordinates of the three phosphors allows specification of the chromaticity map, in an analogous manner to equation (4.1):

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} CC_{mr} & CC_{mg} & CC_{mb} \end{bmatrix} \begin{bmatrix} mr \\ mg \\ mb \end{bmatrix} \qquad (4.4)$$

The chromaticity coordinates of the monitor white point can be found by letting the mrgb coordinates equal (1/3, 1/3, 1/3) in the above equation.

This mapping can be extended to the tristimulus coordinates by introduction of the scaling constants:

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} CC_{mr} & CC_{mg} & CC_{mb} \end{bmatrix} \begin{bmatrix} \rho mR \\ \gamma mG \\ \beta mB \end{bmatrix}$$

The ρ ,γ ,β variables have been moved into the column vector for convenience.

Suppose that the luminance of the monitor white point is L_w. By equation (4.4), the chromaticity coordinates of the white point, (x_w , y_w) can be found. Then by use of the familiar xyY to XYZ transform, the XYZ coordinates of the white point can be found:

$$\begin{bmatrix} x_w / y_w \\ 1 \\ z_w / y_w \end{bmatrix} L_w = \begin{bmatrix} CC_{mr} & CC_{mg} & CC_{mb} \end{bmatrix} \begin{bmatrix} \rho mR \\ \gamma mG \\ \beta mB \end{bmatrix} \qquad (4.5)$$

Since the monitor white point, by definition, has mRGB value of (1,1,1), this system can then be solved for the three unknowns ρ, γ,β.

5. A NOTE CONCERNING HSL TO RGB TRANSFORMATION

5.1. Introduction

Two perceptual spaces are commonly found in computer graphics: HSL space (first defined by the 1979 Core[13]), and *HSV* (first defined by Smith[57]) space.

In the HSL system, a color is defined by its *hue*, *saturation*, and *lightness*. In HSV space (defined in section 3.13.1), a color is defined by its *hue*, *saturation*, and its *value*. The same color will have the same hue in both HSV and HSL space, but (possibly) different saturations.

These are *perceptual* spaces in which the color axes appear to align closely with intuitive human classification of color. For this reason, in many application programs colors are specified using perceptual systems.

Applications often require the coordinates of the color transformed into *RGB* space (discussed in sections 3.6 and 3.9). Many applications (paint programs in particular) require frequent use of these transformations in time-critical situations. For example, the user may wish to change the lightness of an area of the screen, leaving the hue and saturation untouched.

Using *max* to represent the maximum of the R,G, and B components of a color, and *min* to represent their minimum, Lightness (l), value (v), and saturation (s_L , s_V) are defined from RGB values as follows:[13, 27, 57]

$$v = max \qquad \text{(value)}$$

$$s_V = \frac{v-min}{v} \qquad \text{(satv)}$$

$$l = \frac{(min+max)}{2;} \qquad \text{(light)}$$

$$s_L = \begin{cases} \dfrac{max-min}{max+min} & \text{if } l \le \frac{1}{2} \\ \dfrac{max-min}{2-max-min} & \text{if} l > \frac{1}{2} \end{cases} \qquad \text{(satl)}$$

the definition of the hue h in terms of RGB value is irrelevant for present purposes.

This chapter presents an algorithm to invert the operation, determining R, G, and B from h, s_L, and l. This is done by combining portions of two existing algorithms. The first is an HSV to RGB transformation (referred to as HSV_TO_RGB), the second an existing HSL to RGB transformation (referred to as HSL_TO_RGB). These algorithms are reprinted at the end of the chapter.

5.2. The algorithm

Given h, s_L, and l, the algorithm first computes the values of min and max, then the values of s_V and v, and finally the values for R, G, and B.

5.2.1. The computation of max

Consider the case $l \leq \frac{1}{2}$. We attempt to find max.

$$\begin{aligned}
\mathit{max} &= \tfrac{1}{2}(2\mathit{max}) \\
&= \frac{\mathit{min}+\mathit{max}}{2}\left(\frac{2\mathit{max}}{\mathit{min}+\mathit{max}}\right) \\
&= l\left(\frac{\mathit{max}+\mathit{min}+(\mathit{max}-\mathit{min})}{\mathit{min}+\mathit{max}}\right) \\
&= l\left(1 + \frac{\mathit{max}-\mathit{min}}{\mathit{max}+\mathit{min}}\right)
\end{aligned}$$

$$\mathit{max} = l\,(1 + s_L)$$

In the other case, $l > \frac{1}{2}$.

$$\begin{aligned}
\mathit{max} &= \frac{\mathit{min}+\mathit{max}}{2} + \frac{\mathit{max}-\mathit{min}}{2} \\
&= l + \frac{\mathit{min}^2-2\mathit{min}-\mathit{max}^2+2\mathit{max}}{2(2-\mathit{min}-\mathit{max})} \\
&= l + \frac{2(\mathit{max}-\mathit{min})}{2(2-\mathit{min}-\mathit{max})} - \frac{\mathit{max}^2-\mathit{min}^2}{2(2-\mathit{min}-\mathit{max})} \\
&= l + s_L - \frac{(\mathit{max}+\mathit{min})(\mathit{max}-\mathit{min})}{2\,(2-\mathit{min}-\mathit{max})}
\end{aligned}$$

$max = l + s_L - ls_L$

This is an inversion of the derivation performed implicitly by the HSL_TO_RGB algorithm in the computation of variable 'M' (see end of chapter).

5.2.2. Computation of *min*

The derivation of the value of *min* is much simpler:

$$min = min + max - max$$

$$= 2\frac{min+max}{2} - max$$

$$min = 2l - max$$

This is an inversion of the derivation performed implicitly by the HSL_TO_RGB algorithm in the computation of variable 'm' (see end of chapter).

For implementation, a strength reduction[1] is employed:

$$min = l + l - max$$

5.2.3. Computation of v

The computation of v is trivial. By equation (value), $v = max$, which has been determined above. In the remainder of this presentation, *max* will be used interchangeably with v.

5.2.4. Computation of s_V

For the remainder of this algorithm, we assume the color is non-black ($max \neq 0$). This condition is easily detected; when it holds, the algorithm can stop at this point, returning an RGB value of (0,0,0).

For a non-black color, s_V is trivially computed by equation (satv) and the results above:

$$s_V = \frac{max - min}{max}$$

5.2.5. Computation of RGB

Given h, s_V, and v, the computation of R, G, and B can be computed. v and *min* specify the maximum and minimum components of the triple. h determines which of R, G, and B are assigned the minimum, medial, and maximum components. For example, when $h \leq 1/6$, the color is reddish, R receives the maximum component value. s_V and h jointly determine the medial component; the medial component increases with s_V, and decreases as the color approaches a "pure" hue.

The computation is done easily (although opaquely) by the following sequence, produced by a simplification of the end of the HSV_TO_RGB algorithm:

```
real variable stretch, sextant, vsf, mid1, mid2, fract;
stretch := 6 $h$;
sextant := $\lfloor$stretch$\rfloor$;
fract := stretch - sextant;
vsf := $v$  $s_V$ fract;
mid1 := $min$ + vsf;
mid2 := $max$ - vsf;
case sextant of
    0 : ( R, G, B ) := ( $max$, mid1, $min$ );
    1 : ( R, G, B ) := ( mid2, $max$, $min$ );
    2 : ( R, G, B ) := ( $min$, $max$, mid1 );
    3 : ( R, G, B ) := ( $min$, mid2, $max$);
    4 : ( R, G, B ) := ( mid1, $min$, $max$);
    5 : ( R, G, B ) := ( $max$, $min$, min2);
esac;
```

The value of the median component is established through the variable "vsf"; the median value increases with v and s_V, and decreases the closer the hue lies to a pure red, green, blue, cyan, magenta, or yellow (hence the presence of "fract").

5.3. Derivation

The algorithm has been constructed from pieces of two existing algorithms, with an optimization performed in the final computation of R, G, B.

Table 1 : sources for the variable assignments

variable	derived from
max	HSL_TO_RGB
min	HSL_TO_RGB
v	known indirectly
s_V	HSV_TO_RGB
R,G,B	HSV_TO_RGB

5.3.1. Comparison

The standard HSV_TO_RGB and HSL_TO_RGB algorithms, as well as a description of the new algorithm, are given at the end of this chapter.

The following table gives operation counts for the three algorithms. Fractional operation counts for HLS_TO_RGB reflect the fact that the algorithm execution time depends on hue.

	HLS_TO_RGB	new	HSV_TO_RGB
tests	16.33	3	1
calls	3	0	0
floor	0	1	1
:=	9.67	12	9
+	2.67	3	0
-	5	3.5	5
*	2.33	4	6
/	1.33	1	0

For run-time profiling, three algorithms were invoked 1,000,000 times on random input: the standard HLS_TO_RGB, HSV_TO_RGB, and the new algorithm.

At each iteration, the outputs of HLS_TO_RGB and the new algorithm were compared, and any differences reported; none were found.

The programs were written in the C programming language, and executed on a time-shared VAX 11/750 with a floating-point accelerator under the Berkeley Unix operating system. The *gprof* command was used to obtain run-time profiling. The *cc -O* option was used to generate optimized code :

Algorithm	unoptimized		optimized	
	time*	%	time	%
HSV_TO_RGB	176.36	43.30	150.86	46.45
HLS_TO_RGB	407.27	100.0	324.75	100.0
new	232.77	57.15	175.40	54.01

* (CPU seconds)

5.3.2. Conclusion

With a minimum of effort, the HSL_TO_RGB algorithm can be translated into a slight extension of HSV_TO_RGB. This effort was found to reduce HSL conversion time by 43% on random input using unoptimized code, and by 46% on optimized code.

5.3.3. The Standard HSL_TO_RGB Algorithm

From the 1979 Core,[13] page III-38:

```
function VALUE( n1, n2, hue: real ) : real;
begin
   if hue > 360 then hue := hue - 360;
   if hue < 0 then hue := hue + 360;
   if hue < 60 then VALUE := n1 + (n2 - n1)*hue/60;
   else if hue < 180 then VALUE := n2;
   else if hue < 240 then VALUE := n1 + (n2 - n1)*(240-hue)/60;
   else VALUE := n1;
end; { VALUE }

procedure HLS_TO_RGB( h, s, l : real; var r,g,b: real );
begin
   if l <= 0.5 then M := l*(1+s)
            else M := l+s-l*s;
   m := 2*l - M;
   if s = 0 then
      r := g := b := l
   else begin
      r := VALUE( m,M,h+120);
      g := VALUE( m,M,h);
      b := VALUE(m,M,h-120);
   end;
end;
```

5.3.4. The Standard HSV_TO_RGB Algorithm

From Smith.[58]

```
procedure HSV_TO_RGB( h, s, v : real; var r, g, b : real);
var
   f,m,n,k    : real;
   i          : integer;
begin
   h := h * 6;
   i  := h;        /* integer part of hue   */
   f  := h - i;
   m := v*(1-s);
   n := v*(1-(s*f));
   k := v*(1-(s*(1-f)));
   case ( i mod 6 ) of
     0    : (r, g, b) := ( v,k,m);
     1    : (r, g, b) := ( n,v,m);
     2    : (r, g, b) := ( m,v,k);
     3    : (r, g, b) := ( m,n,v);
     4    : (r, g, b) := ( k,m,v);
     5    : (r, g, b) := ( v,m,n);
   end; { case }
end; { HSV_TO_RGB }
```

5.3.5. The new algorithm

```
{ given $h, s_L, l$ on [0..1], return R,G,B on [0..1] }
proc new( in $h, s_L, l$ ; out r,g,b );
local variables
   $v$, $min$, $s_V$;
   vsf;
   sextant;
   fract;
   mid1, mid2;
begin proc
   if ( $l$ <= 0.5) then
      $v := l*(1.0 + s_L)$;
   else
      $v := l + s_L - l*s_L$;
   fi;

   if ( $v = 0$ ) then
      (r,g,b) := (0,0,0);
   else
      $min := l + l - v$;
      { compute saturation in HSV system }
      $s_V := (v - min) / v$;

      $h$   := $h$ *6; { map hue onto [0..6) }
      sextant   := floor($h$);    { integer part of hue   }
      fract     := $h$ - sextant;      { fractional part    }
      vsf   := $v*s_V*$fract ;
      mid1      := $min$ + vsf;
      mid2      := $v$ - vsf;
      case sextant  of
         when 0      : (r,g,b) := ( $v$,mid1,$min$);
         when 1      : (r,g,b) := ( mid2,$v$,$min$);
         when 2      : (r,g,b) := ( $min$,$v$,mid1);
         when 3      : (r,g,b) := ( $min$,mid2,$v$ );
         when 4      : (r,g,b) := ( mid1,$min$,$v$ );
         when 5      : (r,g,b) := ( $v$,$min$,mid2);
      esac;
   fi; { non-black }
end proc
```

6. DISPLAYING THE UNDISPLAYABLE

When sending an mRGB value for display on the monitor, two conditions may arise; the color may be too luminous for monitor reproduction, or it may possess a chromaticity outside the range of the monitor gamut. In this section, two quick and easy algorithms for obtaining "best" approximations are given.

Luckily, such approximations are often nearly indistinguishable from the real color. To quote Hunt,[31]

> when the colour of an object in a colour picture is appraised by an observer, it will generally look acceptable, provided it falls somewhere within the range of colours which that object customarily exhibits in everyday life." - (pp. 43)

6.1. Luminance overflow

Luminance overflow occurs when the computed color has a luminance too great for the monitor. For example, the surface of the sun is too luminous to reproduce accurately. In this case, it is desirable to maintain the chromaticity of the color, solely decreasing the luminance.

Luminance overflow shows when any of the mRGB components are greater than 1. One solution to this overflow is to normalize the vector, dividing each component by the maximum.

While the mRGB system possesses no direct chromaticity dimensions, it can be proven that this mRGB normalization does maintain chromaticity, as a consequence of Grassman's laws.

The chromaticity coordinates mr , mg , mb of the color with luminance overflow (mR, mG, mB) are obtained by

$$mr = \frac{mR}{mR+mG+mB}$$

$$mg = \frac{mG}{mR+mG+mB}$$

$$mb = \frac{mB}{mR+mG+mB}$$

Let t denote the value of the maximum component. Then the chromaticity coordinates mr', mg', mb' of the normalized color (mR/t, mG/t, mB/t) are simply

$$mr' = \frac{\frac{mR}{t}}{\frac{mR}{t}+\frac{mG}{t}+\frac{mB}{t}} = \frac{mR}{mR+mG+mB} = mr \ ,$$

$$mg' = \frac{\frac{mG}{t}}{\frac{mR}{t}+\frac{mG}{t}+\frac{mB}{t}} = \frac{mG}{mR+mG+mB} = mg \ ,$$

$$mb' = \frac{\frac{mB}{t}}{\frac{mR}{t}+\frac{mG}{t}+\frac{mB}{t}} = \frac{mB}{mR+mG+mB} = mb \ ,$$

The chromaticity coordinates are unchanged, the normalized color possesses the same chromaticity as the old.

6.1.1. An obvious solution that fails

Another obvious solution is to "clamp" to 1 those components which exceed unity, leaving the valid components unchanged. However, this can be easily shown to change the chromaticity of the displayed color. Suppose, for example, that the red component did not exceed one. The chromaticity coordinate mr' is derived by

$$mr' = \frac{mR}{mR + max(mG,1) + max(mB,1)}$$

The numerator remains unchanged, but the denominator has changed; obviously, $mr' \neq mr$, the chromaticity has been changed.

This chromaticity change was used by Reeves[54] as part of a quick technique to model the differing chromaticity of fire as its concentration changes.

6.2. Unrealizable color.

The second category of undisplayable color is *unrealizable color*, a color whose chromaticity lies outside the monitor gamut. It is generally agreed that the "best possible" approximations maintain luminance and hue of the unrealizable color, changing solely saturation. The author is aware of two algorithms to display an approximation, and this section presents two new algorithms.

6.2.1. Ives-Abney-Yule

The *Ives-Abney-Yule compromise* is the oldest. The rationale of this compromise is that a "common way in which colours vary in real life is by a uniform addition of white to all colours".[31] Since white lies at the center of the chromaticity diagram, their compromise is to construct an imaginary phosphor triple whose triangle *does* enclose all of color space. The coordinates of a color in this hypothetical color space are then used as the coordinates of the color in the monitor gamut. All colors will therefore be plotted abnormally close to the middle of the phosphor triangle, i.e. perturbed towards white. If only a few colors are unrealizable, this leaves much of the monitor gamut unused.

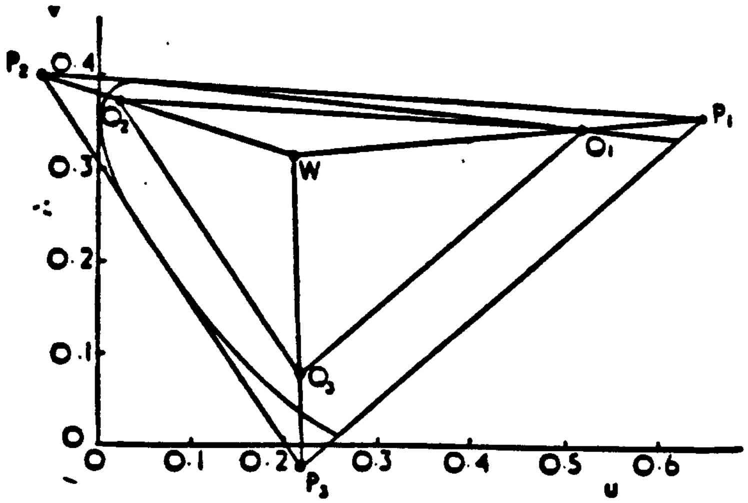

figure 6.1: the Ives-Abney-Yule compromise

from Hunt,[9] pp. 90

Rather than perturb all colors, this perturbation can also be performed only on non-realizable colors. This loses uniqueness of displayed color, but retains accuracy of colors that were realizable beforehand, and lessens the unused portion of the gamut.

6.2.2. Chromaticity clamping in RGB space

In RGB space, this condition manifests when at least 1 component is less than zero. This super-saturation is reflected in the HSV space, the saturation becoming > 1:

$$S = \frac{max-min}{max},$$

$$= 1-\frac{min}{max},$$

$$min<0 \rightarrow 1-\frac{min}{max}>1 \rightarrow S>1$$

By setting the saturation to 1, a color will be created which by definition will be realizable, and by construction differs only in saturation from the original color.

```
Algorithm RGB_Chromaticity_Clamp
    translate the RGB value to HSV
    { the H, V components are OK, S will be > 1.0 }
    set S to 1.0
    translate back to HSV
```

Since the correction is only performed if S > 1, uniqueness of displayed color is lost.

6.2.3. Chromaticity clamping in mRGB space

The preceding algorithm, while extremely simple and provably correct in effect, clamps to *RGB* space, while displayed colors are normally specified in *mRGB* space. As discussed in section (3.9.3), there is a major difference between the two!

Again, chromaticity overflow implies that exactly one of (mR,mG,mB) is less than zero.

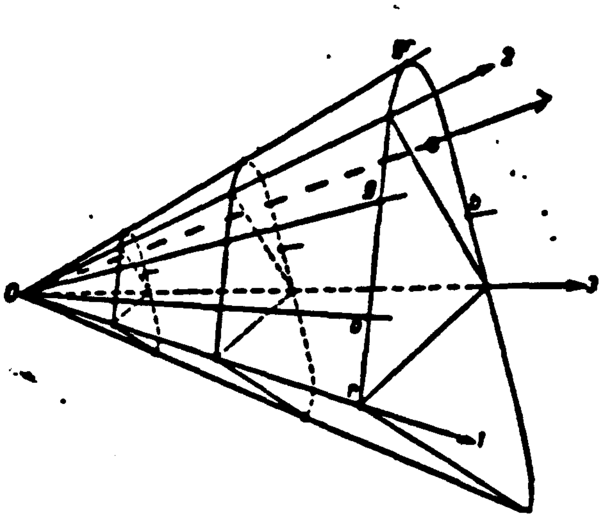

figure 6.2: chromaticity overflow in mRGB space

The algorithm works in mrgb space, by drawing a line between the white point and the color point. The intersection of this line with the monitor gamut represents a color with the same hue as the overflow color, the same luminance, but the maximum possible saturation, just as in the RGB clamping algorithm. The new color point can then be translated back into mRGB space, since the luminance is maintained.

The process, therefore, is to intersect a line drawn from the monitor white point to the undisplayable color with the line of displayable color. The color point at this intersection is then displayed, at the original luminance.

Assume, for example, that mR >= mG >= mB; the other cases follow by symmetry. Then the situation in mrgb space is as below:

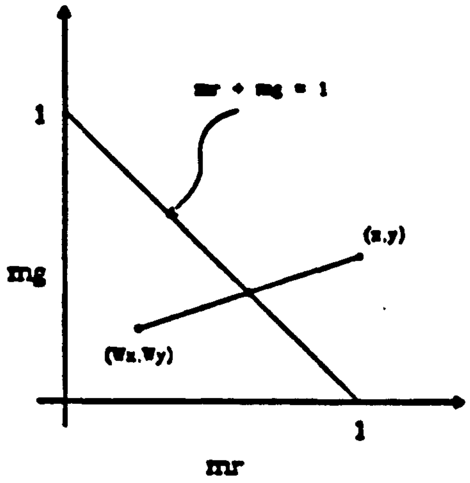

figure 6.3: chromaticity overflow in mrgb space

The line intersects the triangle at the line ($b = 0$). The new color point is then computed by the following procedure:

```
function mRGB_Clamp return point is
    { given a color point (x,y), and white point coordinates (wx, wy),
    intersect the point with the line (x + y = 1)}
    slope := ( y - wy ) / ( x - wx);
    b := wy - slope*wx;
    new_x := ( 1.0 - b ) / ( 1.0 + slope);
    new_y := 1.0 - new_x;
    return ( new_x , new_y );
end
```

Finally, the mrgb coordinates are translated back into mRGB space by multiplying by the luminance.

6.2.4. Cook's algorithm

In 1981, Cook,[14] in a work on specular and diffuse reflection, came upon the same problem. His solution, which is similar to the mRGB clamp, is presented too briefly for detailed comparison. Cook simply states

> "the tristimulus XYZ values are converted to a color space in which locations are specified by dominant wavelength and purity. The purity is then reduced while the dominant wavelength (and thus roughly the hue) is held constant until the color lies inside the monitor gamut" (pp. 19)

The difference lies in the color space of operation. The mRGB clamp works directly in mRGB space, with no need for any other translation. Cook translates from XYZ into $\Lambda\rho$ space, a fairly expensive process, to determine (just as in the author's algorithm) the intersection with the line of displayable color, and then to translate back.

The two algorithms are therefore more alike than they are different. The mRGB clamp, however, is a quicker way of achieving the same effect.

7. TRANSLATING FROM MRGB TO SPECTRAL CURVES

The color algorithms discussed in sections to come all deal with spectral reflectance and transmission curves of color. However, in practice these algorithms will tend to be embedded within graphics applications programs, where the only available information is the mRGB (or sometimes RGB) value of the color.

This section describes a technique to derive a spectral transmission curve from mRGB value, and a "quick and dirty" approximation technique to derive a spectral reflectance curve from mRGB value.

The transformation from a spectral curve to a mRGB triple maps from an infinite-dimensional space (the set of curves) to a three-dimensional one (the set of triples). This many-to-one mapping results in the phenomenon of *metamerism*; there are an infinite number of spectral curves that will yield the given mRGB value.

Therefore, what is desired is simply to construct one possible curve, a considerably easier task.

This can be done easily by use of one piece of information; the spectral transmission curves of the red, green, and blue phosphors of the monitor being used. These can either be measured directly, or can often be obtained (with slightly less accuracy) from the manufacturer.

Each of the three curves can be easily stored in computer memory, in discrete array representation.

Using Grassmans second law, the transformation is then extremely simple. Assume the mRGB value is (ρ, γ, β). Furthermore, let $Rmap(\lambda)$, $Gmap(\lambda)$, and $Bmap(\lambda)$ represent the spectral transmissions of the primaries. Then the transmission $T(\lambda)$ of the color (ρ, γ, β) is given by

$$R(\lambda) = \rho Rmap(\lambda) + \gamma Gmap(\lambda) + \beta Bmap(\lambda) \tag{7.1}$$

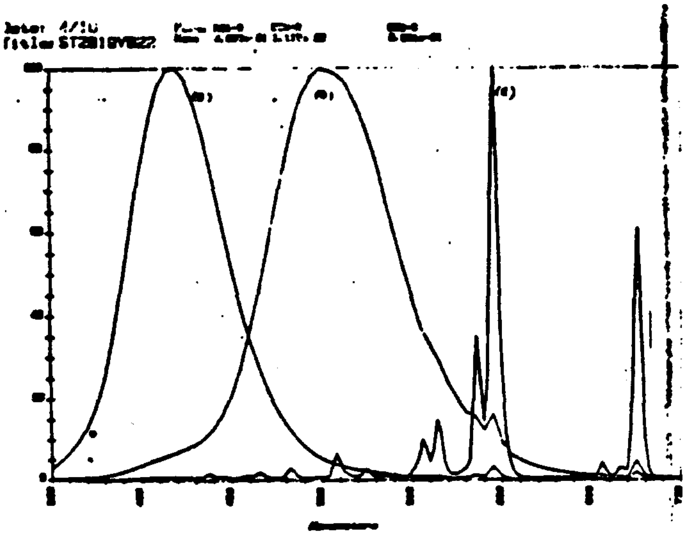

figure 7.1: spectral information for a typical monitor

the red, green, and blue curves are scaled by .41, 1.11, and .28.

7.1. mRGB to reflectance curve

When a reflectance curve is desired, the situation is similar. The transmission curve obtained by the above equation can be considered an approximation to the *reflectance* curve; a height of 0.5 at 540 nm can be considered either a transmission of 0.5 units of energy (transmission curve) or as a reflectance of 50% of the incoming light (reflectance curve).

7.2. Limitations

This quick approximation has two disadvantages. First, as mentioned above, it returns *a* curve, not necessarily *the* curve desired. Second, the monitor curves represent *transmission* characteristics, while they are being used here to create *reflection* curves. This approximation is often perfectly valid; a green light shining off a white paper is the same as to a white light shining off a green paper. The only danger arises if any component of R is > 1, a perfectly valid situation in terms of transmission curves, but a semantically meaningless one for reflectance curves. In this case scaling must be done.

8. Additive Color Mixture

8.1. Introduction

This chapter is the first of three dealing with *color mixture*, a term referring to the different techniques by which color is changed by physical processes; mixing of paints, combination of light, shining light through filters, etc.

There is disagreement about not only the laws followed by the different types of mixture, but even the basic question of the *number* of types of mixture! Answers range from two[27, 51] to four[42] to eight.[20] In this presentation, four types of color mixture will be considered, a subset of the eight types mentioned by Duncan:[20] *additive*, *filter subtractive*, *dye subtractive*, and *pigmentary*.

This chapter considers the first type of color mixture, *additive mixture*.

8.2. Additive mixture

When two or more lights are shone on each other, the color of the resulting light is determined by the principles of *additive mixture*. The resulting color is formed by the sum of the transmittances of the incoming lights, hence the term 'additive':

$$T_{out}(\lambda) = \sum_i c_i T_i(\lambda) \qquad (8.1)$$

where

$T_{out}(\lambda)$ is the transmission of the output light at wavelength λ,

c_i is the concentration of the *i'th* light, and

$T_i(\lambda)$ is the transmission of the $i'th$ input light at wavelength λ.

8.3. Additive mixture in RGB space

Since RGB space is additive (due to Grassman's second and third laws), additive mixture can be simulated very easily in RGB space. Given the presence of n colored lights in absolute concentrations $c_1 \cdots c_n$, the color of the resulting light is simply

$$(RGB)_{out} = \sum_{i=1}^{n} c_i (RGB)_i \tag{8.2}$$

The equation is modified slightly in the case of additive mixture of surface colors. If small dots of one color are interspersed with small dots of another, the resultant color perceived is the average rather than the sum of the constituent colors:

$$(RGB)_{out} = \frac{\sum_{i=1}^{n} c_i (RGB)_i}{\sum_{i=1}^{n} c_i} . \tag{8.3}$$

This equation forms the basis for all anti-aliasing algorithms, as well as the artistic technique of pointilism.

9. SUBTRACTIVE MIXTURE

9.1. Introduction

When colored light is shone through one or more filters, or transmitted through a colored liquid, the color of the transmitted light is determined by the laws of ***subtractive mixture***. Each filter(surface) absorbs a certain percentage of the incoming light, and transmits(reflects) the remainder. process is used in color photography, filtering, and inking.

This chapter deals first with the case of light shining through filters, and then with the case of light shining upon a colored dye.

9.2. Filtration

In classical subtractive mixture, a colored light is shone through one (or more) colored filters. The theory determining the color of the transmitted light is fairly simple, and was first determined by Bouger, Beer, and Lambert over 200 years ago.

Each filter absorbs a certain percentage of the incoming light flux at each wavelength, and transmits the remainder. No backwards reflection is assumed.

Suppose the incoming light has a transmittance curve T, and the n filters have transmittance curves of $T_1, T_{2,} \cdots T_n$ per unit thickness. Then the transmittance curve of the resultant light U is simply

$$U = T \prod_{i=1}^{n} T_i \tag{9.1}$$

There is a slight difference in the semantics of the transmittance curves. The curve for the incoming light, T, refers to absolute light flux present at each wavelength; the filter curves refer to the *ratio* of the transmitted flux to the penetrating flux, the so-called "internal spectral transmittance".[39]

Filters with different thicknesses can be accommodated easily, as a filter of thickness X behaves identically to X filters of unit thickness. If, for example, a filter with transmittance T_1 of thickness 3 is laid on top of a filter with transmittance T_2 of thickness 2, then the transmittance is

$$U = T \, T_1^3 \, T_2^2$$

This is simply a special case of equation (9.1).

9.2.1. Implementation

Given equation (9.1), the question now becomes one of implementation: Can subtractive mixture be simulated given only the RGB (or mRGB or XYZ) values of the incoming light and the filters?

Of course, due to metamerism, an infinite number of spectral curves can derive the same RGB values; it is well-documented that different filters with the same RGB value can have wildly different visual performance.[5, 20, 23, 39, 63, 37] Therefore, the problem is really to

> "[find] a function that yields plausible results. A 'plausible' result is one which would be expected based on experience based on experience with ordinary colored materials..." - Joblove,[37] pp. 35

The author is aware of one existing algorithm to model subtractive mixture in the literature using heuristics, and two other algorithms are presented.

9.2.1.1. Joblove

In his Masters' thesis,[37] Joblove attempted to model subtractive mixture by performing an algebraic transformation into a new color space, termed "ij chromaticity space", and performing geometric interpolation in that space. Joblove since found the algorithm to be inaccurate, and has discarded it.[35]

9.2.1.2. The RGB Heuristic

The RGB values of the filters can be considered as an approximation to the "percentage transmittance" of the color in each of the broad areas of the spectrum. For example, an orange filter (RGB (1,.5,0)) transmits all the reddish, half the greenish, and none of the blue light that lands on it. Using this heuristic, Filtration mixture is easily simulated as[25]

$$(R,G,B)_{out} = \left(\prod_{i=0}^{n} R_i \ , \ \prod_{i=0}^{n} G_i \ , \ \prod_{i=0}^{n} B_i\right) \tag{9.2}$$

For example, passing cyan light through an orange, and then a grey filter produces:

cyan	- (0 , 1 , 1)
orange	- (1 , .5 , 0)
grey	- (.5 , .5 , .5)
result	- (0 , .25 , 0)

The author implemented this program in 1982 at the University of Wisconsin, where it was used by Meteorologists interested in remote sensing. While this algorithm is a gross simplification of the actual physics demonstrated in equation (9.1), it seemed to meet the users' requirement for "plausible results".

9.2.1.3. mRGB heuristic

Chapter (7) gave an algorithm to obtain transmittance curves from mRGB values. Using this method, transmittance curves can be obtained for each of the filters and the incoming light, and equation (9.1) followed precisely:

```
Algorithm mRGB_Heuristic
compute transmittance curve approximation of each filter
for each wavelength do
    take the product of all transmittances by equation (9.1)
od;
use equation (7.1) to reconstruct mRGB from new curve
```

This algorithm has been implemented, and some of its results are shown on the following color page. It has been the authors' experience that this algorithm very well simulates filtration mixture. Particularly, it is the only algorithm of the four that simulates the impurities found in actual filters.

9.3. Dyes

The second case of classical subtractive mixture concerns the mixture of dyes. For present purposes, a dye may be defined as a liquid or *soluble* solid that imparts its color to any material immersed in the dye.[28] There are two essential differences between dyes and pigments, the subjects of the next chapter. First, dyes *dissolve completely* into their medium, behaving as a *solution*. This is in marked contrast to pigments, which remain undissolved *in suspension*. Second, (due in large part to the first difference), *dyes do not scatter light*, while *pigments may*.[21,46] When light encounters a particle of dye, it is either absorbed by the dye or transmitted through it (undergoing classical subtractive filtration in the transmission). In the case of pigments, *all three* phenomenon may occur.

In order to simplify the treatment of pigments and dyes, the light is assumed perpendicular to the surface, an assumption made by most in the field.[39]

The classic Fresnel equation for reflectance from a beam perpendicular to the surface is

$$\rho = \left(\frac{n_2 - n_1}{n_2 + n_1} \right)^2 , \tag{9.3}$$

Where ρ is the Fresnel reflectance, n_2 the index of refraction of the optically denser medium and n_1 that of the optically less dense. In the case of common oils n_2 = 1.5, in the case of air n=1.0, and ρ evaluates to only 0.04. In general, index of refraction may be a function of wavelength,[23] a case that will not be considered here.

Of course, this is not the sole factor in the answer; light may be multiply reflected *internally*, and finally this *internally diffuse* light may be emitted.

The physics are relatively simple. As shown in figure (9.1) below, portions of the light are either reflected or transmitted at each layer boundary. It is assumed that the background lies "in front of, but not optically attached to"[39] the dye layer. For this reason, this simulation only applies to colored substances such as glass, gelatine, plastic, varnish, and lacquer film. The optics involved when the background optically contacts the colorant are discussed in section (9.3.2).

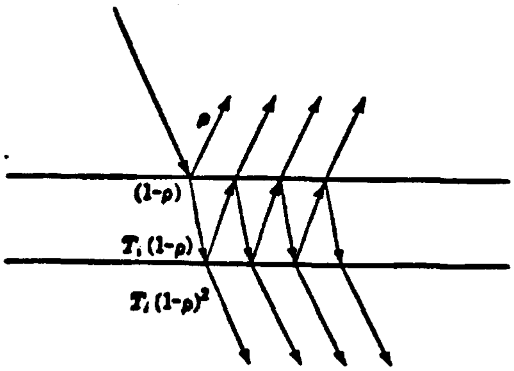

figure 9.1: dye transmission and reflection

from Judd,[39] *pp. 331*

In the present case, the percentage reflected is[39]

$$R = \rho + \rho(1-\rho)^2 T_i^2 (1+\rho^2 T_i^2+\rho^4 T_i^4 \cdots) \qquad (9.4)$$
$$= \rho + \frac{\rho(1-\rho)^2 T_i^2}{1-\rho^2 T_i^2}$$

In the case of filtered mixture, each filter was represented as a separate component in the product. In the case of dyes, only *one* layer exists, the layer between the surface and the background. There are four variables in dye mixture: the *thickness* X of the layer of dye, the *absorption* K of the dye, the *concentration* c of the dye, and the *transmission* T of the dye. In the simple case of dyes, when no scatter occurs, these quantities are related by Beer's and Bouger's laws,

$$T = e^{-cKX} \qquad (9.5)$$

The new variable K represents the percentage of incoming light absorbed by the dye per unit thickness. K is a function of wavelength for each dye, not a single constant. Due to the absence of scatter, the transmission can thus be found solely in terms of the absorption and the thickness.

The above equation only supplies one value for K. In the case of *dye mixture*, when two or more dyes are present, it was first shown by Duncan[19] that the absorption coefficient of the mixture is the weighted average of the absorption coefficients of the dyes. In the case where the medium of suspension itself absorbs or colors the light (as is the case, for example, in heavy varnishes and tinted oils), the medium can be considered as simply another weight in the computation of K.[19]

9.3.1. Implementation

To review, the percentage of light reflected from the surface of a dyed surface (assuming incident light perpendicular to the surface, and a background not in optical contact with the layer of colorant) can be determined precisely, and is a function of the index of refraction of the medium, the concentration of the dye, the thickness of the layer, and the absorption coefficient(s) of the dye(s).

```
Algorithm Dye_Mixture
evaluate $\rho$ by equation (9.3)
for each wavelength do
    $K_M$ := weighted average of all absorptions for this wavelength
    compute $T_i$ by equation (9.5)
    then compute $R$ by equation (9.4)
od
compute RGB from series of $R$'s computed
```

Again, from the point of view of Computer Graphics, the problem becomes more difficult. The indices of refractions and absorption coefficients must become known to the graphics algorithm, from (possibly) nothing more than the mRGB values of the dyes.

Varying simplifications can be made to obtain values for these variables:

1) The surrounding medium can be assumed as air, with index of refraction of 1.0003.[39]

2) The index of refraction of the medium can be assumed known by user specification. Failing this, as most media used as vehicles have indices of refraction ranging from 1.48 (light oil) to 1.66 (heavy oil), the program could simply pick a representative value. 1.5 may be a particularly good choice, lying extremely close to the indices for both glass, rosin, soybean oil, linseed oil, and shellac.

3) The concentration can either be set to 1 (Bouger's case), or determined by user specification as part of the interactive process.

4) Since the absorption coefficients are a function of wavelength, it is impractical and unfair to assume user specification. This question is still very much an open one. The author is currently investigating a number of heuristics, including

    4.1) Using the technique outlined in chapter (7) to derive transmission curves of the dye, and (due to the absence of scatter), letting $K_i = 1-T_i$.

    4.2) An even cruder, but much quicker, approximation is to use the RGB value as an approximation, for each third of the spectrum, of the transmission. For example, consider a dye formed by yellow and red colorants. Then the dye would be assumed to transmit perfectly in the red, transmit 50% in the green, and not transmit at all in the blue. This simple RGB interpolation can be performed, (within the other assumptions!) due to the interpolatory nature of absorption coefficients.

9.3.2. A final note

In the preceding section, the background lay behind, but did not optically contact, the colorant layer. This is unfortunate from a graphics programming point of view; the ability to vary the background color, and see the concommitant change in the colorant appearance (in the case of little concentrated or thinly layered dye) is desirable.

In this case the mathematics become slightly more complex. Updating figure (9.1) for the new case, the figure becomes

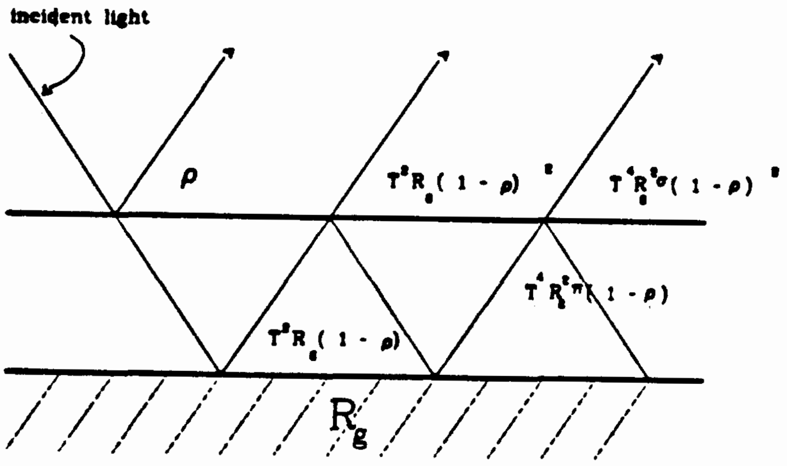

figure 9.2: a dyed colorant against a background

In this case, by resorting to the above figure, it can be seen that the reflectance R is

$$R = \rho + (1-\rho)^2 \sum_{i=1}^{\infty} T^{2ci} R_g{}^i \rho^{i-1} \tag{9.6}$$

Where R_g is the reflectance of the background. Since ρ is known less than ½ and each of the other terms are guaranteed ≤1, this sum may be converted into closed form:

$$R = \rho + \frac{(1-\rho)^2}{\rho} \sum_{i=1}^{\infty} (T^{2c} R_g \rho)^i$$

$$R = \rho + \frac{(1-\rho)^2}{\rho(1-T^{2c} R_g \rho)} \tag{9.7}$$

10. PIGMENTARY MIXTURE

When two or more colored pigments are mixed in a suspension, the color of the suspension is determined by the process of *pigment mixture*. This process is by far the most complex of the types of mixture discussed.

Paint programs, by definition, are often used to simulate the process of painting. Accordingly, it is desirable for paint programs to contain algorithmic simulations of the paint mixture process. This chapter first presents the optics of paint mixture, with a brief description of the different equations that have been developed. We conclude that the physics of paint mixture are too complex for a computer program to realistically and quickly implement; we then present a heuristic algorithm to simulate "ideal" paint mixture for a much simpler set of cases.

10.1. Terminology

Just as in subtractive mixture, consider a light shining upon a surface. The intensity of the light will be denoted as I. The surface paint, also called the *colorant*, consists of one or more different types of *pigment particles*, which remain *in suspension* in the medium of suspension, termed the *vehicle* or *binder*. The paint rests upon a background, termed the *canvas* or *ground*. If the paint is sufficiently thick so that the canvas plays no role in the perceived color of the surface, the paint is said to be of *hiding thickness*. The degree to which a given paint of a given thickness obscures the canvas is termed *hiding power*.

The physics of the situation are quite complex, and are shown pictorially in the figure below:

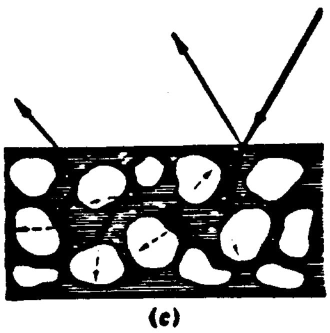

figure 10.1: the path of light through paint

from Evans, pp. 280

When the incident light contacts the surface of the paint, some portion is reflected from the surface immediately by the Fresnel equation of section (9.3). Some is transmitted into the paint, and some is *scattered*, changing direction within the paint.

Once within the paint, a ray of light may contact a particle of pigment. Upon doing so, the process is repeated: some of the light is absorbed, some transmitted, and some reflected. Light within the paint is termed *internal diffuse* light. The amount of internal diffuse light that eventually is transmitted *back* across the surface of the paint is termed the *internal diffuse reflection*.

The total reflectance from the paint is the *sum*, therefore, of the surface reflectance ρ and the *diffuse reflectance* R. The amount of diffuse light that is transmitted through the paint is termed T, similarly to subtractive mixture.

Paint optics are therefore more difficult than dye optics, due to the complex interaction between the light and the pigment. Yet it will be shown that even this optics model is a drastic simplification.

Pigment mixture uses both additive and subtractive mixture in its modeling. Additive mixture occurs when the diffuse internal reflected light, consisting of a number of rays of different colors, is combined with the light reflected by the surface. Each time light contacts a pigment particle, or is transmitted or absorbed, classic subtractive mixture occurs.

10.2. The Physics of Paint

This section gives an extremely brief description of the optical behavior of paint. The cardinal equations and results are introduced in historical order, with a brief mention of the assumptions and limitations of each. Liberties have been taken with the notation used by the particular papers, in order to ensure a uniform notation:

Notation for Paint Optics	
symbol	meaning
I	incident flux
R	diffuse reflectance
R_i	diffuse reflectance at distance i from the surface
R_g	reflectance of the ground
ρ	incident reflectance
T	diffuse transmittance
T_i	diffuse transmittance at depth i from the surface
K	coefficient of absorption;
	the percentage of flux absorbed per unit thickness
S	coefficient of scatter;
	the percentage of flux scattered per unit thickness
γ	$\sqrt{K(K+2S)}$
X	thickness of the paint
dx	thickness of a layer of paint
y	amount of flux moving towards the ground from the surface
z	amount of flux moving towards the surface from the ground
φ	the Duncan reflectivity function
c	concentration
c_i	concentration of the i'th pigment

10.2.1. Kubelka & Munk (1931)

Consider an infinitely thin slice of the paint, of thickness dx. Due to light scattering caused by pigment particles, light can enter the layer from *both* the top and bottom. A certain flux I enters the layer from some direction. By definition, $K\,I\,dx$ of the flux will be absorbed, and $S\,I\,dx$ will be *scattered*. The remainder, $I-(K+S)\,I\,dx$ will be transmitted through the layer. This creates the following system of differential equations, the *Kubelka-Munk equations*:[41]

$$dy = -(K+S)\,y\,dx + S\,z\,dx \;, \tag{10.1}$$

$$dz = (K+S)\,z\,dx - S\,z\,dx \;,$$

Where y represents the flux moving from the top to the bottom, and z the flux moving in the opposite direction. y_i refers to the flux at a thickness i from the surface.

In these terms, the percentage of diffuse reflectance R is $\left(z_0 \,/\, y_0\right)$, the ratio of reflected light to incident light. Similarly, the diffuse transmittance T is $\left(y_X / y_0\right)$ for a paint of thickness X.

10.2.2. Amy (1937)

Amy[2] was the first to apply the Kubelka-Munk equations to paint mixture. Amy solved the Kubelka-Munk differential equations and simplified, yielding

$$T = \frac{(2\gamma e^{-X\gamma})}{\psi} \;, \tag{10.2}$$

$$R = \frac{S\left(1-e^{-2X\gamma}\right)}{\psi} \;,$$

where

$$\gamma = \sqrt{K(K+2S)} \;,$$

$$\psi = K+S+\gamma-\left\{K+S-\gamma\right\}e^{-2X\gamma}$$

Amy then simplifies these equations greatly by adding two constraints. Firstly, if $K^2 \gg S^2$, then $\gamma \approx K+S$ and $\psi \approx 2(K+S)$. The equations become

$$T = e^{-(K+S)X} \quad , \tag{10.3}$$

$$R = \frac{S\left\{1-e^{-2X\gamma}\right\}}{2(K+S)} \quad ,$$

Secondly, if at least one of X,S, or K is large, then $e^{-X\gamma} \approx 0$. This constraint yields a simplified reflectance of

$$R = \frac{S}{2(K+S)} \tag{10.4}$$

For example, when $K=1/6$, $S=1/3$, this evaluates to .333.

This simplified set of equations exhibits two desirable properties: the diffuse transmittance T varies inversely with the exponential of the thickness, and the diffuse reflectance R varies inversely with the coefficient of absorption K.

Amy's equations, while simple, make three large assumptions: one of (K , S , X) is large, $K^2 \gg S^2$, and the reflection from the canvas can be ignored.

10.2.3. Duncan (1940)

The major paper in the field of pigment mixture is that of Duncan.[19] Duncan uses the Amy formulae as "a basis from which, making certain assumptions, it is possible to deduce an equation by means of which the colours of paints containing a mixture of different pigments may be predicted correctly".[19]

Amy's third assumption was that canvas reflection could be ignored. This is the "certain assumption" kept by Duncan, and its consequences expanded. If the canvas reflectance can be ignored, the paint must be of hiding thickness, by definition. In this case "the thickness is such that any further increase therein produces no further change in colour by reflected light".[19] By referring to equation (10.2), it follows that $e^{-2X\gamma} = 1$, and equation (10.2) becomes

$$T = \frac{2\gamma}{K+S+\gamma}\,, \qquad (10.5)$$

$$R = \frac{S}{K+S+\gamma}$$

Amy's formula made two assumptions involving K and S. *This* formula represents the general case, "in which K and S may have any values and the proportion and nature of the pigments added may be such as to cause considerable variation in both K and S".[19]

Both equations (10.5) and (10.4) are derived from the same complex equation, (10.2); the equations are simplified differently depending on the assumptions made.

The equation for R above may be re-written as

$$\varphi = \frac{S}{K} = \frac{2R}{1-R^2} \qquad (10.6)$$

φ is the so-called "reflectivity function",[19] which essentially characterizes the entire problem in terms of *one* parameter.

Equation (10.6) can be derived by solving equation (10.5) as a quadratic in S, the other root being $S = 0$. $S = 0$ is precisely the case of dyes, a special case of pigmentary mixture in which R evaluates to 0, there being no internally reflected light.

Duncan solves the problem of pigment mixture quickly by making two assumptions: that the absorption and scatter coefficients of the mixture are the weighted averages of the absorption and scatter coefficients of the individual pigments:

$$\varphi_M = \frac{S_M}{K_M} = \frac{c_1 S_1 + c_2 S_2 + \cdots c_n S_n}{c_1 K_1 + c_2 K_2 + \cdots c_n K_n} \tag{10.7}$$

Duncan provides no theoretical justification for this equation; it is assumed, and then confirmed by empirical testing.

As a quick example, suppose two pigments are mixed in a 2:1 ratio. Consider the computation of R for some particular wavelength, say 614 nm. Assume furthermore that $K_1 = .1$, $K_2 = .3$, $S_1 = .4$, and $S_2 = .2$. Then the average $K = .167$, the average $S = .333$. Then by equation (10.6), $\varphi = \frac{.333}{1.67} = 2$. R can then be indirectly approximated by use of tables[19,39] as roughly .38.

This result could also be oomputed by direct evaluation of equation (10.5):

$$R = \frac{.333}{.167 + .333 + \sqrt{.167(.167 + 2(.333))}}\ , \ = .38$$

There are two signal contributions in Duncan's paper: the derivation of an equation (10.5), which describes the reflectance of a paint without any restrictions on the K and S values of the paint, and the "center of gravity" rule for the K and S values of a mixture of pigments.

The Duncan algorithm has several limitations:

1) The paint is assumed of hiding thickness.

2) The medium of suspension is assumed clear and colorless, with index of refraction equal to that of the pigments (which are all assumed to possess equal indices).

3) The K and S values hold only for a certain medium. if the same pigments are suspended in a different medium, the values must be re-established experimentally; no transformation algorithm exists.[19,20,21]

10.2.4. Duntley (1942)

The most complex set of equations dealing with diffusing materials are those of Duntley.[22] Duntley attempted to remove all assumptions regarding the optical properties of the incident light and the material. The cost of this increased generality is increased computational complexity, and the use of *eight* constants, as opposed to the two of Kubelka and Munk, Amy, and Duncan.

A detailed presentation is beyond the scope of this paper. Essentially, Duntley notes that the internally diffused light operates under different optical properties than the incident light, and has different scattering and absorption coefficients. In addition, the percentage of the light which is scattered backwards versus the-percentage scattered forwards "are, in general, not the same for the incoming light as for the internal diffuse light". Finally, the flux moving towards the surface undergoes slightly different refraction from that moving towards the background.

With these additional factors taken into account, With these relaxed assumptions, the differential system becomes (using Kubelka-Munk notation for the variables involved)

$$dI'_x / dx = -(K' + B' + F')I'_x \ ,$$

$$dz / dx = F'I'_x - Kz - Bz + By \ ,$$

$$-dy / dx = B'I'_x - Ky - By + Bz \ ,$$

Where

I'_x is the unscattered residue of the primary beam at depth x,

K is the absorption coefficient for internal diffused flux,

K' is the absorption coefficient for incident flux,

S is the total scattering coefficient for internal diffused flux,

S' is the total scattering coefficient for incident light,

B is the percentage of S scattered backwards,

B' is the percentage of S' scattered backwards, and

F is the percentage of S scattered forwards,

F' is the percentage of S' scattered forwards.

This system is solved, yielding the following equations for diffuse transmittance T and diffuse reflectance R:

$$T = \frac{Q'x_i + P'e^{-(K'+B'+F')X} B \sinh x_i X}{(K+B)\sinh x_i X + x_i \cosh x_i X} - (Q'-1)e^{-(K'+B'+F')X} \,, \qquad (10.8)$$

$$R = \frac{P'x_i e^{-(K'+B'+F')X} + Q'B \sinh x_i X}{(K+B)\sinh x_i X + x_i \cosh x_i X} - P' \,,$$

where

$$x_i = \left[K(K+2B)\right]^{1/2} \,,$$

$$P' = \frac{(K-K')B' + (B-B')(B'+F')}{(K'^2-K^2) + 2K'(F'+B') - 2KB + (F'+B')^2} \,,$$

$$Q' = \frac{(K+K')F' + (B+F')(B'+F')}{(K'^2-K^2) + 2K'(F'+B') - 2KB + (F'+B')^2}$$

The above equations are daunting indeed. Duntley shows that in the specific case of dyes (no scatter, $B=B'=F=F'=0$) the solutions reduce to those of the preceding chapter, $R = 0$ and $T = e^{-KX}$. Similarly, in the case of infinite thickness, the solutions become $T=0$ and

$$R_\infty = \frac{Q'B}{K+B+\gamma} - P'$$

The last simplification that can be performed is to assume that $P'=0, Q'=1$, in which case Duntley states his equations "represent a two-constant theory", the special case referred to previously.

Finally, the Fresnel equation for surface reflection at the surface boundary was described in section (9.3) as $\rho = (\frac{n_2-n_1}{n_2+n_1})^2$, for light with normal incidence. In the case of diffuse light, with the simplifying assumption that the light is traveling in air, Duntley cites the much more complex equation of

$$\rho = \tfrac{1}{2} + \frac{(n-1)(3n+1)}{6n+1^2} + \left[\frac{n^2(n^2-1)^2}{(n^2+1)^3}\right]\log\left(\frac{n-1}{n+1}\right) - \frac{2n^3(n^2+2n-1)}{(n^2+1)(n^4-1)} + \left[\frac{8n^4(n^4+1)}{(n^2+1)(n^4-1)^2}\right]\log n \qquad (10.9)$$

Duntley's extremely complex equations pose a number of severe problems for computer implementation. First, these equations need to be evaluated across a spectrum, not simply once; the computational cost alone renders the technique impractical for real-time computation. Second, and more severely, the formula requires use of *eight* constants, *each* of which is a function of wavelength. Duntley recommends that these constants be derived experimentally for each particular mixture, which is obviously infeasible for computer graphics. No tables of these constants are available†, making computer implementation nearly impossible.

Finally, it can be seen that even a mild increase in the complexity of the incident light behavior (diffuse vs. normal incidence) leads to a large increase in computational complexity.

10.2.5. Saunderson (1942)

Saunderson[55] found Duntley's equations slow and inconvenient, and uses the two-constant theory of Duncan, with minor optimizations made for the case of pigmented plastics.

Saunderson's main contribution was the introduction of a correction factor: the value of the diffuse reflectance *R* "corresponds to

† to the author's knowledge

measurements made with the specimen immersed in a liquid of the same index of refraction".[39] In order to correct for this,§ Saunderson showed that the true reflectance R' can be obtained from R by

$$R' = \rho + \frac{(1-\rho)(1-\rho_i)R}{(1-\rho_i R)} , \qquad (10.10)$$

Where ρ_i is the internal reflectance for completely diffuse light incident on the underside of the top layer. In general, this must be determined experimentally; for plastics, values between 0.4 and 0.6 are typical.[18,39,55]

10.2.6. Kubelka (1948)

In 1948, Kubelka[40] derived a system of *explicit* solutions for his 1931 system of equations, with background reflectance specifically included. In this system, R can be computed directly as

$$R = \frac{1-R_g(a-b \coth b\ S\ X)}{a-R_g+b \coth b\ S\ X} , \qquad (10.11)$$

$$a = \frac{(S+K)}{S}$$

$$b = \sqrt{a^2-1}$$

This equation is perhaps the most useful of the four presented for R ((10.1),(10.2),(10.5),(10.11)), giving R directly in terms of every variable of interest.

For paint of hiding thickness, the above becomes

$$R = \frac{1}{a+b \coth bSX}$$

If the paint is of hiding thickness, the thickness can be increased arbitrarily. For infinite thickness, the equation further simplifies to

§ paintings being rarely viewed underwater

$$R = \frac{1}{a+b},$$
$$= \frac{S}{S+K+\sqrt{K^2+2KS}}$$

Which is precisely the special case of equation (10.5).

10.2.7. Duncan (1949)

In 1949, Duncan[20] published a second paper on pigmentary mixture, which restates and expands on his first. Duncan makes several points in this paper that aid in computer implementation:

1) In his original paper,[19] Duncan mentioned that the K and S values held only for the medium of measurement. In this paper, Duncan expands the point:

> "The values of the coefficients of scatter, and, to a lesser extent, of the coefficients of absorption of a pigment vary with the refractive index of the medium in which it is used. Values obtained in one medium therefore cannot be used for paints in another medium unless it is of approximately the same refractive index... In general terms, however, it can stated that the coefficient of scatter falls as the refractive index of the medium approaches that of the pigment. The coefficient of absorption also falls slightly, at least in some cases, owing presumably to reduction in the amount of internal reflection inside the individual particles of pigment" pp. 303

2) Duncan states that the correction factor of Saunderson[55] (equation (10.10)) has only a minor effect, and can be ignored in practice. However, in a later paper[21] Duncan reverses this conclusion.

3) The accuracy of the computation need not be high:

> "For colour prediction work the coefficients of scatter and absorption of the pigments of interest should be determined for rough work at a minimum of 7

> wavelengths scattered throughout the spectrum; it would be advantageous to take about twice this number for greater accuracy... there is little point in striving at an accuracy exceeding 5-10 per cent in the values of [K and S] or two significant figures in the prediction of reflectances" - pp. 305

10.3. Discussion

Four equations for diffuse reflectance have been presented; those of Amy, Duncan, Duntley, and Kubelka.

The equation of Amy is unsuitable for general use, due to the restricted range of K and S for which it holds. Specifically, for white (highly scattering) paints, the equation will predict significantly less reflectance than correct.

The equation of Duntley, while perhaps the superior in terms of realism, requires six more constants than any other the others, which can only be derived experimentally.

Duncan's equation simply reduces to a special case of Kubelka's 1948 equation. Therefore, it appears that the optimal choice is as follows: if the user desires interaction with the canvas color, then the full Kubelka-Munk equation (10.12) is preferable. If the canvas can be ignored, then the simplified Kubelka-Munk equation of Duncan (10.5) can be used.

The equation for surface reflectance must also be chosen; the Fresnel equation (9.3), or Duntley's more complex (10.9). Unfortunately, the Fresnel equation cannot be chosen out of hand; as the next figure demonstrates, the two reflectances can be drastically different:

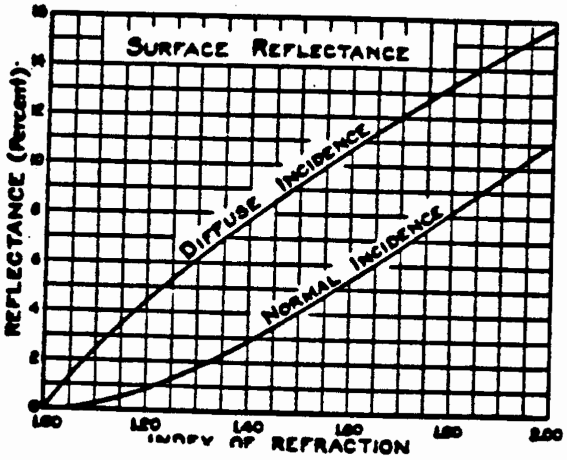

figure 10.2: surface reflectance for diffuse and normally incident light from Duntley,[22] pp. 66

For a typical oil paint, with index of refraction 1.5, the normally incident light has 4% surface reflectance, and the diffuse 9%.

Due to this difference, this choice should be left to the user; in the case of user indifference, the quicker Fresnel equation could be used.

10.3.1. A pseudo-code implementation

The results can be summarized by the following procedure to simulate paint mixture:

```
proc Paint_Mixture
    for each wavelength $\lambda$ do
        $I$ := incident flux at wavelength $\lambda$
        $\rho$ := surface reflectance , medium index of refraction $n$
            (by either equation (9.3) (10.9) )
```

```
          $R$ := diffuse reflectance ,
               evaluated by result of Choose_Mixture_Equation
          $R'$ := $R$ corrected by equation (10.10)
     od
     the $R'$ curve holds the reflected color
end proc
```

This algorithm, though slow*, is easily implementable, with one large exception; the provision of values for the variables:

n) The index of refraction of the medium, while in general a function of wavelength, can be assumed a constant for most purposes. Just as in subtractive mixture, this can be either entered by the user or any number of tables can be consulted for representative values.

I) The intensity of incident flux is a function of wavelength. Using the technique of chapter (7), only the mRGB (or, in this case, even RGB) coordinates need be entered. Again, this simplicity makes user specification attractive. Otherwise a "typical" soft white could be used.

R_g) The reflectance of the background can be approximated using the mRGB heuristic of chapter (7). If the background is unimportant to the user, then equation (10.5) can be used, an equation where no specification of background reflectance is needed.

K,S)The K and S values are the weighted averages of the K and S values of the component pigments. The problem therefore becomes to find the K and S values of each pigment, given only their RGB (or mRGB) values. Even if only their English names are known to artistic users,

*In the simplest case (using equation (9.3), Duncan mixture, and only one pigment), the algorithm requires 5n adds, 4n subtracts, 8n multiplies, 3n divides, and n square roots, where n is the number of wavelengths sampled.

references (albeit obscure)[3] exist giving xyY or XYZ values of common pigments.

This is a very difficult problem indeed; the K and S are a function of medium, and are always determined experimentally in practice. Heuristics to create reasonable values are currently under study.

10.3.2. Limitations of the Theories

More recent tests have shown that some of the basic equations may be limited in applicability. While most tests have verified the linear relationship between K/S and concentration, recently Hattori[30] found this relationship did not hold for very high wavelengths, while Beresford[4] found a more severe inaccuracy due to imperfections in paint and dye technology.

Perhaps the greatest limitation lies in the K and S values for a given pigment. As explained in section (10.2.7), these values are accurate only for *a given medium*; the K and S values for a pigment in oil, for example, can be drastically different from those of the same pigment in water. This is highly unfortunate, and hinders severely the creation of a model that accurately handles changes in medium.

10.3.3. Assumptions Common to the Theories

The preceding formulae, while of substantial complexity, nevertheless make a number of simplifying assumptions in order to ease the mathematics:

1) *particle interaction*. All presentations have assumed no chemical or electrical interaction between either pigments of differing types, or pigments and medium. In fact, such interaction occurs in virtually all paintings, and often has a dramatic impact on the color of the

paint[12, 43, 17, 28] In fact, an entire literature dealing solely with the chemical interactions between paints in a vehicle exists, with a theory all its own.

2) *particle clumping*. Even ignoring chemical interaction, it is assumed that the particle grains are uniformly and evenly distributed throughout the medium. In practice, some degree of *flocculation* (clumping) occurs in most paints, (indeed, some amount of flocculation is a desired goal in paint preparation)[17, 28] and, again, can drastically change the color which the paint would present normally.[17, 65]

3) *particle shape*. All presentations have assumed that the light transmitted through and reflected by the particles of pigment is diffuse in all directions. This is only true if the pigment particles are spherical. In practice, even ignoring flocculation, pigment particles tend to be acircular, taking cylindrical, bullet-like, or teardrop-shaped form.[24, 60, 48]

4) *particle boundary interaction*. When light leaves a pigment particle and re-enters the medium, refraction, diffraction, and transmission all occur, just as when light enters the medium from the scene.[48] The light traverses a boundary between medium of two different refractive indices. This effect has been ignored by all presentations. According to Judd,[39] the effect can be ignored only if the suspended particles have dimensions less than 10% the wavelength of the incident light, a condition which Duntley[22] specifically states not to hold for pigments. Introducing this effect introduces another physical theory all its own, with truly horrendous computational difficulty. For one thing, index of refraction of pigments tends to vary with wavelength;[48] modeling the refraction alone is extremely complex (see for example Pangonis).[52]

5) *particle size*. In addition to assuming that all particles are spherical, they are assumed to be of the same size, a gross simplification.[20,48,60]

6) *the canvas shape*. For those presentations that deal with the case when the paint is not of hiding thickness, the canvas is sometimes assumed to be a perfect reflector, and *always* assumed to be perfectly smooth. In practice, neither is the case; even the whitest canvases reflect only 80% of incoming light, and rough canvas is, in fact, *required* in order to hold the paint to the canvas, in the case of watercolor and pastel paints.[46,47,43]

7) *the surface shape*. Similarly, the surface the paint presents to the external environment is assumed to be of perfect smoothness. While this assumption is reasonable for pastel and watercolor, it is a gross distortion for tempera and especially oil. Much of the appearance of oil paintings is specifically due to the ridges and furrows built up in its surface.[46,47,7]

8) *particle floating*. The treatments have assumed the pigment particles perfectly dispersed throughout the paint. There are a number of cases where this assumption does not hold: in mat or semi-mat paint films, where air penetrates into the top layers,[39] in "paint floating" where the pigment grains of higher fineness float towards the top,[12,43,39] and in paint settling and drying, where a quite complex process of settling, drifting, and opacity change occurs.[12,43,65,46]

10.3.4. Applications to other fields

Throughout this chapter, the topic of paint mixture has been discussed. This reduces to (in large part) a discussion of the interaction of light with particles suspended in a medium.

This is exactly the situation encountered in two other areas of interest in computer graphics, the simulation of Clouds and of Fog.[21] In both cases, particles (of water) are in suspension (in air), particles which partially absorb, partially reflect, and partially transmit the incoming light. Indeed, the very first paper of interest in simulation of paint mixture[56] was mainly interested in the visibility of objects through fogs.

The treatments of both fog and clouds possess several large simplifications compared with pigmentary mixture; there is (generally) only *one* type of particle, which *is* spherical, which *is* evenly distributed, and which *has* well-known refractive index and color. In addition, in the case of fog, absorption can be generally ignored.[61] Of course, this also allows the introduction of another entire body of work and research to be brought to bear (see for example[61,62,29,9,45,50]).

There are two complicating factors. First, clouds and fog are not (in general) static media. Over a short time-frame, they may be considered so, but in general their boundaries, size, and density all vary with time. Second, and more importantly, the geometric boundary is no longer as well-behaved. In the case of paint, a horizontal sheet of paint lay on a plane of canvas, illuminated only from above. In the case of fog and clouds, a major part of the simulation lies in determination of the irregular and shifting boundary of the medium, and the diffuse nature of the illumination.

10.3.5. Conclusion

The degree to which the faulty assumptions and limitations weaken the power of the equations is extremely difficult to judge. It appears that in industry the equations tend to be used more as a guide than as a firm arbiter,[4,6,16,30,53] Kubelka's algorithm being within 3% of the actual value over 67% of the spectrum in a typical example.[4]

All applications, furthermore, were performed in environments in which K and S could be established by testing for each pigment. When these values are not readily available, the problem becomes even more difficult.

However, just as in subtractive mixture, what is really desired is not *the* result, as is the case in industry, but rather *a plausible* result, a result a user of a computer graphics program would accept. The author has developed one such heuristic algorithm, and another is currently under development.

10.4. A paint mixture heuristic

This section presents a heuristic algorithm designed to approximate a user's intuitive expectations of paint mixture.

Referencing rules of common mixtures found in art textbooks,[6,11,33,34,44] the author derived five "plausibility criteria" a heuristic algorithm should obey:

1) *any paint mixed with itself does not change color.*

2) *any paint mixed with black maintains chromaticity and becomes darker.*

3) *any paint mixed with white maintains chromaticity and becomes lighter.*

4) *any color mixed in equal proportions with its complementary produces grey.*

5) *blue paint and yellow paint mix to green paint.*

Of course, the literature is rife with pathological cases where one or more of these rules fail abysmally. However, these rules appeared both to the author and to the artistic references consulted to represent "normal" mixtures.

The fourth rule may appear circular, as a paint's complement is often *defined* as the color of paint that, when mixed with it, produces grey paint. However, there are a number of "standard" complements that can be referenced:

Blue and Orange

Red and Green

Yellow and Purple

The fifth rule is debatable, due to the substantial ambiguity surrounding the word "blue". Some claim this refers to RGB (0 , 0 , 1) , others that it refers to RGB (0 , 1 , 1) ("cyan"). The author used the former interpretation.

The algorithm maps each input paint color to a point in some color space. The color point of the resultant paint is the center of gravity of the input points, where the weights correspond to the relative proportions of the paints in the mixture.

This center of gravity law serves to satisfy the first criterion, regardless of color space; the center of gravity of a point is itself.

The center of gravity law (all mixtures of two paints lie on a line between them) in combination with the second criterion (mixing with black darkens, and maintains chromaticity) implies that the line between any color point and the black point is a line of constant chromaticity. This condition holds in either HSL, HSV, or RGB space.

Similarly, the third criterion (mixing with white lightens, and maintains chromaticity) implies that the line between any color point and the white point is a line of constant chromaticity. This condition holds *only* in HSL space; in HSV space, moving towards white will lighten and desaturate a color, changing its chromaticity.

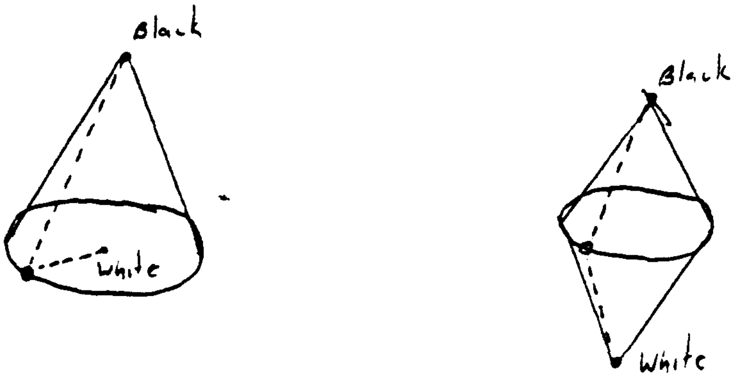

figure 10.3: mixing with black and white in HSV and HSL

The first three rules, therefore, imply that the color space of operation be HSL, rather than HSV space.

The arrangement of hues around the circle is somewhat arbitrary; any arrangement which preserves isomorphism is valid. The fourth and fifth criteria can be satisfied by choosing a hue arrangement with red, yellow, and blue at 120 degree angles to each other, a *RYB* hue wheel:

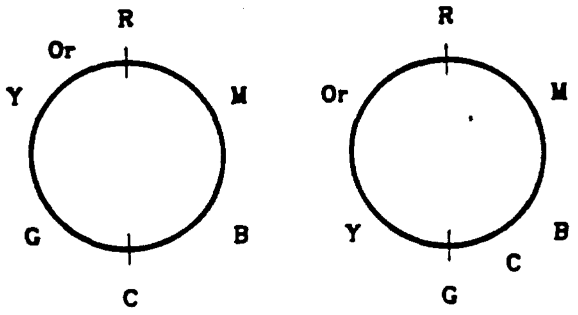

figure 10.4: RGB and RYB hue wheels

The two wheels are related by a simple mapping. This HSL space with the RYB hue wheel will be denoted RYB_HSL space. The paint mixture heuristic algorithm can be written simply as

```
{ given a list P of paint colors in RGB, and a list W of relative weights,
return color of mixture in RGB }
func Paint_Mixture_Heuristic : returns RGB value
    let Q := a list of every color in P, translated into RYB_HSL space
    let C := center of gravity of Q, weighted by W
    let r := C, translated from RYB_HSL to RGB
return ( r )
end func
```

For implementation, the RGB to RYB_HSL conversion can be accomplished by using the standard RGB to HSL conversion, and changing the HSL hue into the RYB_HSL hue:

```
func Hue_Map ( hsl_hue : real ) : return real
    if hsl_hue ≤ 1 / 6 then
        return 2*hsl_hue { stretch yellow }
    elseif hsl_hue ≤ 2 / 3 then
        return ( ( ( hsl_hue - 1/6 ) / 2 ) - ( 1 / 3 ) ) { squeeze green }
    else
        return hsl_hue
    fi
```

end func

The author first implemented this algorithm in January 1982 at the University of Wisconsin,[25] and it has been at use there since that date. It has also been installed in the Berkeley paint program[26] since March of 1983. In both cases it appears to have met users criteria for plausible results.

10.4.1. Disadvantages

The pigmentary heuristic has one severe disadvantage; *it is inflexible.* Blue and yellow will always mix to the same shade of green. There is no room in the model for simulation of media, of canvas, of lighting, or of drying.

11. CONCLUSION

"He shoots ... he scores !"

Color is the medium of communication of raster computer graphics; every image, every algorithm, and every scene rely on color science.

Applications of color science span the gamut of computer graphics. In this thesis, a number of applications have been presented, resulting in new algorithms for computer graphics:

1) A new derivation of a standard color system translation, resulting in an algorithm (chapter (5)) nearly twice the speed of the existing algorithm.

2) Two algorithms to quickly and optimally display approximations to colors which are not reproducible on the display device (chapter (6)).

3) Two algorithms to approximate the color resulting from the passage of light through filters or off surfaces (chapter (9)).

4) A heuristic algorithm for the simulation of pigment mixture.

for number sequence only

for number sequence only

References

1. Alfred V. Aho and Jeffrey D. Ullman, *Principles of Compiler Design*, Addison-Wesley, Reading, Massachusetts (1977). third edition, 1979.

2. Lucien Amy, "Sur la Coleur des Corps par Reflexion," *Revue d'Optique, Theorique, et Instrumentale* **16**(3) pp. 7-92 (March, 1937). translated for the author by John Gross.

3. Norman F. Barnes, "A Spectrophotometric Study of Artists' Pigments," *Technical Studies in the Field of the Fine Arts* **7**(3) pp. 120-138 (January, 1939). also appears in *Journal of the Optical Society of America*, **29**, 1939, pp. 208-214.

4. J. Beresford, "Instrumental measurement of colour," *Journal of the Oil and Colour Chemists Association* **53** pp. 800-820 (1970).

5. Wilhelm Von Bezold, *The Theory of Color in its relation to Art and Art-Industry*, L. Prang and Company, Boston (1876).

6. Faber Birren, *History of Color in Painting*, Van Nostrand Reinhold Company, New York, New York (1965).

7. D. M. Bishop, "Brushability and Viscosity Control," *Paint Technology* **33**(4) pp. 16-20 (April, 1969).

8. W. R. Blevin and W. J. Brown, "Total Reflectances of Opaque Diffusers," *Journal of the Optical Society of America* **52**(11) pp. 1250-1255 (November, 1962).

9. James F. Blinn, "Light Reflection Functions for Simulation of Clouds and Dusty Surfaces," pp. 21-29 in *Proceedings of SIGGRAPH '82*, ACM,(July, 1982).

10. Pieter Johannes Bouma, *Physical Aspects of Colour*, N. V. Phillips Gloeilampenfabrieken, Eindhoven, the Netherlands (1947). A. Wolters-Dawson, trans.

11. Louis Cheskin, *Colors : What they can do for you*, Liveright Publishing Company (1947).

12. Arthur H. Church, *The Chemistry of Paints and Painting*, Seeley, Service & Company, London (1915). Fourth edition.

13. Graphics Standards Planning Committee, "Status Report of the Graphics Standards Committee," *Computer Graphics* **13**(3)(August, 1979).

14. Robert L. Cook and Kenneth Torrance, "A Reflectance Model for Computer Graphics," pp. 307-316 in *Proceedings of SIGGRAPH '81*, ACM,(August 1981).

15. T. N. Cornsweet, *Visual Perception*, Academic Press, New York (1970).

16. William B. Cowan, "An Inexpensive Scheme for Calibration of a Colour Monitor in Terms of CIE Standard Coordinates," pp. 315-321 in *Proceedings of SIGGRAPH '83*, ACM,(July, 1983).

17. Frederick K. Daniel, "Flocculation and Related Pigment Interactions in Paints," *Paint Technology* **29**(2) pp. 14-23 (February, 1965).

18. Hugh R. Davidson and Henry Hemmendinger, "Color Prediction Using the Two-Constant Turbid-Media Theory," *Journal of the Optical Society of America* **56**(4) pp. 1102-1109 (August, 1966).

19. D. R. Duncan, "The Colour of Pigment Mixtures," pp. 380-390 in *Proceedings of the Physical Society* , London(1940).

20. D. R. Duncan, "The Colour of Pigment Mixtures," *The Journal of Oil and Colour Chemists Association* **32**(7) pp. 296-321 (July, 1949).

21. D. R. Duncan, "The Identification and Estimation of Pigments in Pigmented Compositions by Reflectance Spectrophotometry," *Journal of the Oil and Colour Chemists Association* **45**(5) pp. 300-324 (May, 1962).

22. Seibert Q. Duntley, "The Optical Properties of Diffusing Materials," *Journal of the Optical Society of America* **32**(2) pp. 61-70 (February, 1942).

23. Ralph M. Evans, *An Introduction to Color*, John Wiley and Sons, Inc., New York (1948).

24. B. Felder, "Uber die Tailchengrossenabhangigkeit der Lichtabsorption in heterogenen Systemen," *Helvetica Chemica Acta* **49** pp. 440-453 (1966).

25. Ken Fishkin, "Color Mixture in Computer Graphics," *Transactions of the Wisconsin Academy of Science* **71**(2)(December, 1983). in process.

26. Ken Fishkin, *Blot Users Manual*, University of California, Berkeley, California (1983). Berkeley Computer Graphics Laboratory internal document.

27. James D. Foley and Andries van Dam, *Fundamentals of Interactive Computer Graphics*, Addison-Wesley Publishing Company (1982).

28. Noel Heaton, *Outlines of Paint Technology*, J. B. Lippincott Company (1940). second edition

29. Theodore Edward Heindsmann, *An Investigation of the Transmission of Visible Light Through Artificial Water Fogs*, Master's Thesis, University of California, Berkeley, California (1948).

30. Tadashi Hittori, Kenji Shirai, Miki Niwa, and Yuichi Murakami, "Quantitative Interpetation of Infrared Diffuse Reflectance Spectra over Whole Concentration Range," *Bulletin of the Chemical Society of Japan* **54**(7) pp. 1964-1967 (1981).

31. R. W. G. Hunt, *The Reproduction of Colour*, John Wiley and Sons, Inc., New York (1975). Third edition.

32. Richard S. Hunter, *The Measurement of Appearance*, Hunter Associates Laboratory, Fairfax, Virginia (1973).

33. Michel Jacobs, *The Art of Colour*, Doubleday, Page and Company (1924).

34. Michel Jacobs, *The Study of Colour*, Doubleday, Page and Company (1927).

35. George Joblove, Private communication. July 1983.

36. George H. Joblove and Donald P. Greenberg, "Color Spaces for Computer Graphics," pp. 20-25 in *Proceedings of SIGGRAPH '78*, ACM,(August, 1978).

37. George H. Joblove, *Color Space and Computer Graphics*, Master's Thesis, Cornell University, Ithaca, N.Y. (January, 1979).

38. Deane B. Judd, "Introduction to Color," pp. 1-14 in *Symposium on Color - it's Specification and use in Evaluating the Appearance of Materials*, American Society for Testing Materials,(March 5, 1941).

39. Deane B. Judd, *Color in Business, Science and Industry*, John Wiley and Sons, Inc., New York (1952).

40. Paul Kubelka, "New Contributions to the Optics of Intensely Light-scattering Materials," *Journal of the Optical Society of America* **38** p. 448 (1948). Part I.

41. P. Kubelka and F. Munk, "Ein Beitrag zur Optik der Farbanstriche," *Zurich tech. Physik*. **12** p. 593 (1931).

42. Harald Kuppers, *The Basic Law of Color Theory*, Barron's, New York (1982). Roger Marcinik, trans.

43. A. P. Laurie, *The Painter's Methods and Materials*, Seeley, Service & Co., London (1960).

44. M. Luckiesh, *Color and its Applications*, Van Nostrand Reinhold Company, New York (1927).

45. Nelson Max, John Blunden, and Don Vickers, *Geometric Forms for Clouds and Smoke*, Report, Lawrence Livermore National Laboratory (1983). unpublished.

46. Ralph Mayer, *The Artist's Handbook of Materials and Techniques*, Viking Press, New York (1941).

47. Ralph Mayer, *The Painter's Craft*, Penguin, New York (1979). Third edition.

48. H. E. Merwin, "Optical Properites and Theory of Color of Pigments and Paints," pp. 494-530 in *Proceedings of the 20th Annual Meeting of the American Society for Testing Materials*, American Society for Testing Materials,(1917).

49. Gary W. Meyer and Donald P. Greenberg, "Perceptual Color Spaces for Computer Graphics," pp. 254-261 in *Proceedings of SIGGRAPH '80*, ACM,(July, 1980).

50. M. Minnaert, *The Nature of Light and Colour in the Open Air*, Dover, New York (1954).

51. Roy Osborne, *Lights and Pigments: Colour Principles for Artists*, John Murray, Ltd., London (1980).

52. William J. Pangonis, Wilfried Heller, and Arvid Jacobson, *Tables of Light Scattering Functions for Spherical Particles*, Wayne State University Press, Detroit (1957).

53. Allan M. Raff, "Measurement of a Color Gamut Obtained from FD&C Colorants," *Journal of Pharmaceutical Sciences* **53**(4) pp. 380-385 (April, 1964).

54. William T. Reeves, "Particle Systems - A Technique for Modeling a Class of Fuzzy Objects," pp. 359-376 in *Proceedings of SIGGRAPH '83*, (July, 1983). Also in *ACM Transactions on Graphics*, **2**, 2, pp. 91-108.

55. J. L. Saunderson, "Calculation of the Color of Pigmented Plastics," *Journal of the Optical Society of America* **32**(12) pp. 727-736 (December, 1942).

56. Arthur Schuster, "Radiation Through a Foggy Atmosphere," *Astrophysical Journal* **21**(1) pp. 1-22 (January, 1905).

57. Alvy Ray Smith, "Color Gamut Transform Pairs," pp. 12-19 in *Proceedings of SIGGRAPH '78*, ACM,(August, 1978).

58. Alvy Ray Smith, *Color Tutorial Notes*, Report No. 37, LucasFilm (1981).

59. Alvy Ray Smith, *Painting Tutorial Notes*, Report No. 38, LucasFilm (1982).

60. N. P. Suryanarayana, Mohammad Yaseen, and J. S. Aggarwal, "Critical Pigment Volume Concentration of Paints," *Paint Technology* **30**(6) pp. 28-32 (June, 1966).

61. Herman Thal-Larsen, *Experimental Investigations of Light Transmission and Reflection in Fog*, Master's Thesis, Univerisity of California, Berkeley, Calfironia (1939).

62. Wells Thompson, *The Selective Scattering of Colored Light in Fog*, Master's Thesis, University of California, Berkeley, California (1938).

63. W. D. Wright, *The Measurement of Colour*, Macmillan, New York (1958). second edition

64. G. Wyszecki and W. S. Stiles, *Color Science*, John Wiley and Sons, Inc., New York (1967).

65. Rudiger Ziegelitz, "The Settling of Pigments in Binders," *Paint Technology* **30**(2) pp. 23-26 (February, 1966).