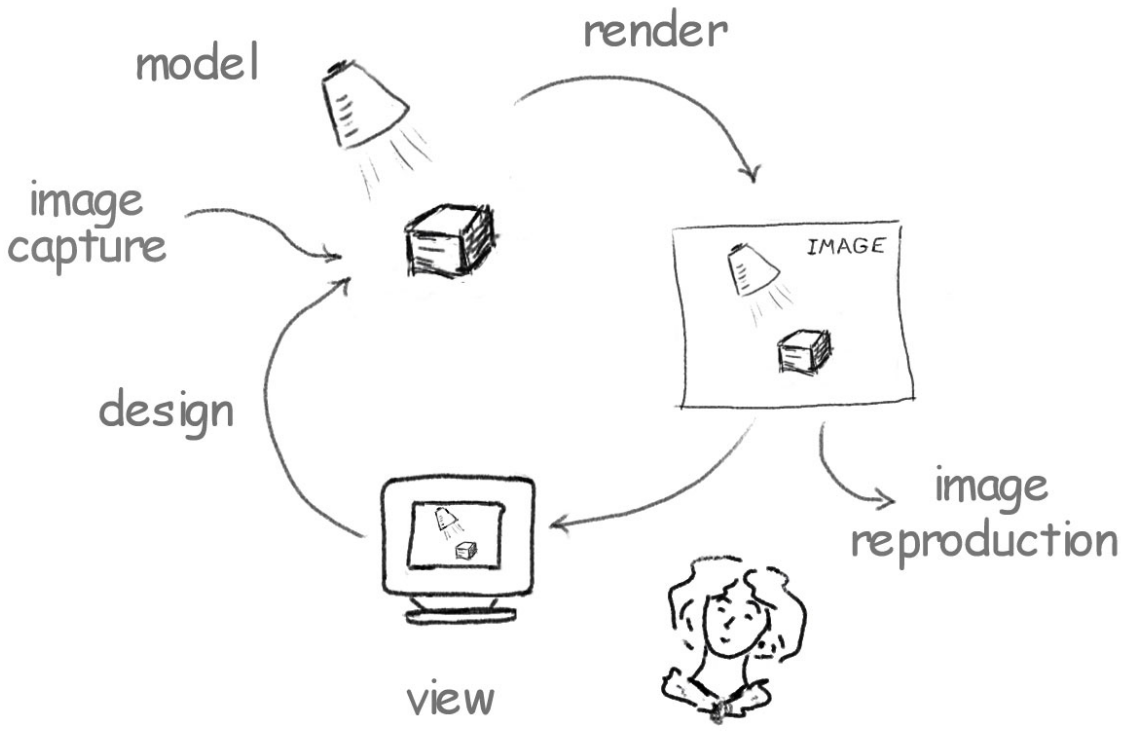
A Field Guide to Digital Color

Course 21
SIGGRAPH 2002
July 22, 2002



Maureen Stone
stone@stonesc.com
StoneSoup Consulting
http://www.stonesc.com
650-559-9280

Pauline Ts'o
tso@rhythm.com
Rhythm & Hues
http://www.rhythm.com
310-448-7500

Description

This course presents a guide to the color disciplines relevant to Computer Graphics ranging from color vision to color design. Participants should leave with a clear overview of the world of digital color, plus pointers to in-depth references on a wide range of color topics. The course should be of value both to those seeking an introduction to color in computer graphics and to those knowledgeable in some aspects of color who wish to get a broader view of the field.

The first part of the course is a survey, presented by Maureen Stone. For each topic, she will provide a brief overview and pointers to further information. For most topics, a standard reference book or two will be suggested, augmented by published articles as needed.

The final hour of this course will be presented by Pauline Ts'o of Rhythm & Hues. She will describe how color is designed and managed in the production of computer generated imagery for film and video.

Presenters

Maureen Stone is an independent consultant working in the areas of digital color, large format displays, web graphics, interaction and design. Before founding StoneSoup Consulting, she spent 20 years at the Xerox Palo Alto Research Center where she attained the position of Principal Scientist. She has over 30 published papers plus 12 patents on topics including digital color, user interface technology and computer graphics. She has taught in SIGGRAPH digital color courses organized by William Cowan (SIGGRAPH '89) and Charles Poynton (SIGGRAPH '97), and first presented this survey course at SIGGRAPH '99. She has a BS and MS in Electrical Engineering from the University of Illinois, and a MS in Computer Science from Caltech. Her book, "A Field Guide to Digital Color" is scheduled for publication fall 2002.

A computer programmer by training, with degrees from Cornell and UC-Berkeley, Pauline Ts'o started her artistic education by rummaging through her mother's art book collection as a child. Her technical training got her a job as a software designer at Robert Abel & Associates, a pioneer in the computer animation industry. From there, Pauline and five co-workers founded Rhythm & Hues Studios in 1987. Her stealth art education came in handy when she began to work as a lighting technical director on many projects including the first Coca-Cola "Polar Bear" commercial. Her most recent credit is as a lighting director for "Harry Potter and the Sorcerer's Stone". In her spare time, Pauline also co-leads photography workshops throughout the Southwest and is working on her first children's book.

Schedule

Module 1: Modeling, Perceiving and Reproducing Color	
8:30	Vision and Appearance - Stone
9:15	Color Reproduction and Management - Stone
9:45	Color Models: Natural and Simulated – Stone
10:15	Break

Module 2: Color in Graphics and Design	
10:30	Color in Graphics Systems - Stone
10:45	Color Selection and Design - Stone
11:15	Color at Rhythm & Hues - Ts'o
12:15	Lunch

Table of Contents

Introduction and Goals

The goal of this course is to provide a broad survey of the color disciplines that relate to computer graphics. After taking this course, you will hopefully know what there is to learn, and where to learn about it. If some of the fundamental principles can be learned just from the presentation, that would be great also.

The survey part of this course is a slightly expanded version of the tutorial presented in previous years called: A Survey of Color for Computer Graphics. This year, the course has been expanded to include a specific example of the way these principles are applied in computer graphics. Pauline Ts'o will describe how color is used in the production of computer generated imagery for film and video at Rhythm & Hues.

These notes are a description of the material that will be presented in the course. They are split by presenter, with the survey material first, and the description of color at Rhythm & Hues second. The slides for the survey part of the course are placed at the end of the notes.

A Field Guide to Digital Color: Survey

Maureen Stone
StoneSoup Consulting

Introduction

Color is an enormous field, with applications in almost every area of life. The taxonomy I present in this survey is designed to match common applications in graphics. It is, by definition, incomplete and superficial. Therefore, I base my presentation on standard texts and references in the field of color as much as possible.

The survey part of this course is divided into three areas: vision, reproduction and synthesis. The vision section provides the foundations for understanding color and its applications. The reproduction section discusses color media and color management. The synthesis section is split into computer graphics rendering, color selection and color design. Included is an extensive annotated bibliography for these topics, plus printouts of the slides.

A more complete version of the material in these notes is in the process of becoming a book, *A Field Guide to Digital Color*, to be published by A.K. Peters, hopefully by the fall of 2002.

Color Vision

In this section, I will discuss basic color vision, colorimetery and color appearance. The reference I use for color vision is Brian Wandell's *Foundations of Vision*, which is structured in three parts: Encoding, representation[1] and response. What I call "basic color vision" is essentially the encoding part of his taxonomy. Colorimetery uses these principles to measure color. Wyszechki and Stiles *Color Science* is considered the standard reference for colorimetery, though most of what is presented here can be found in almost any color text. Color appearance is the response of the visual system to color. Mark Fairchild's *Color Appearance Models* is a recent book on color appearance that includes both a complete description of color appearance as it is currently understood, and a survey of computation models that are being developed to address this complex topic.

Basics

Light enters the eye as spectrum of colors, distributed by wavelength. This *spectral distribution function* impinges on the retina in the back of the eye and is absorbed by the *cones*. Human beings have three types of cones, which respond to different wavelengths of light. These are called either long, medium and short wavelength cones, or, correspondingly, red, green and blue cones. Each cone absorbs light and sends a signal to the brain. That is, the spectrum of light is encoded into three values that correspond to the

[1] Brian's representation section focuses on the visual pathways in the brain—I will not address these issues at all.

amount of light absorbed by each type of cone. This is the principle of *trichromacy*—human vision is 3 dimensional.

It should be clear that the actual distribution of the spectrum is only indirectly "seen" by the eye. The response of each cone can be encoded as a function of wavelength (the *spectral response curve* for the cone). Multiplying the spectrum by such a function and integrating produces the signal that is sent from the eye to the brain. Different spectra can produce the same signal, and as such, will "look the same." This principle is called *metamerism*.

The principle of metamerism underlies all color reproduction technologies. Instead of reproducing the spectral stimulus, they create an equivalent response, or *metameric match* by mixing primary colors in a controlled way.

Trichromacy and metamerism can also be applied to the problem of measuring color. It is important in many industries to be able to measure colored materials. If we can create an instrument that responds in the same way as the human eye, we can have an impartial observer to define when colors match. From the discussion above, it seems obvious to fit an instrument with filters and sensors that behave like the cones. However, the precise definition of the cone response was not know until very recently. The science of color measurement is much older, and is based on experiments involving matching colors with sets of three primary lights. This is called *colorimetery*.

Colorimetery

The color matching experiments that underlie colorimetery are constructed as follows. Choose three primary lights (call them red, green and blue). Then, take a set of reference colors such as the monochromatic colors of the spectrum, or those generated by filtering a white light. For any set of three independent primaries, you can match any other color. Period. The only requirement is that you allow the primaries to go negative. In a physical matching experiment, this means that you shine the "negative" primary on the sample to be matched. This is a remarkably useful result. It means we can create a representation for color based on three primary lights. Any color can be defined as a set of three numbers that correspond to the power of the primaries. These three numbers are called the *tristimulus values* for the color.

How do we know what primary values match a particular color? Clearly we cannot perform the color matching experiment on all possible spectra. The answer is to use the fact that colored light is additive. A spectrum can be constructed by adding a number of monochromatic lights together. This seems obvious, but what is less obvious is that the amount of the primaries needed to match colored light is also additive. Specifically, let the primaries be R, G and B. Let the spectrum be S. If RGB_1 matches S_1, and RGB_2 matches S_2, then RGB_1+RGB_2 will match S_1+S_2. This principle was first formalized by Grassman, it is called *Grassman's law*.

To define the primaries for an arbitrary color, we use Grassman's law to construct a set of *color matching functions* for the primaries. We perform the color matching experiment on each of the monochromatic spectral colors (sampled every 2 nm, for example). We use the result to create a set of three functions, one for each primary. To determine the amount of the primaries needed to match a particular spectrum, multiply the spectrum by the color matching functions and integrate. Three multiplications and three integrations, and you have the match. As all the functions described are sampled, this process is quite simple to implement.

What if we decide to use a different set of primaries? Do we have to repeat the color matching experiments? Most wonderfully, the answer is no, we can compute the color matching functions for the new set of primaries from the old ones. All we need to know is the definition of the new primaries in terms of the old ones. For those of you familiar with linear algebra, this is simply a change of basis, and can accomplished for a three dimensional system with a 3x3 transformation matrix. You apply this matrix to each of the samples in the color matching functions in turn to create the color matching functions for the new set of primaries.

The color matching functions are remarkably like the spectral response curves for the cones. Multiply each by the viewed spectrum, and integrate to get an encoding into three numbers. Are these two types of curves related? The answer is yes, that the cone response, too, is linearly related to the set of color matching functions. This result was only recently demonstrated by Bayer, Nunn and Schnapf in 1987.

In 1931, the *Commission International de l'Eclairage (CIE)* standardized a set of primaries and color matching functions that are the basis for most color measurement instruments used today. They transformed a set of color matching functions measured by Stiles and Burch to create a set of curves that were more convenient to use. This set is positive throughout the entire visible spectrum, and one of the curves can be used to compute the perceived brightness of the measured color. The CIE standard tristimulus values are notated X,Y and Z. They are often reduced to two dimensions by projecting them onto the X+Y+Z=1 plane, creating the CIE *chromaticity diagram* with its corresponding *chromaticity coordinates, x* and *y*.

Colored objects can be described as having spectral reflectance functions that selectively reflect and absorb light shined on them. A "red" object reflects the red wavelengths of light and absorbs the rest. Colorimetery can be applied directly to the light reflected from objects. It is sometime applied to the reflectance spectrum (measured by reflecting a known light off of a sample) as well. However, this spectrum cannot be perceived directly. What is viewed is always a function of both the object and the light.

Some of the basic concepts of vision are demonstrated in the *Color Playground*, created at Brown University (http://www.cs.brown.edu/exploratory/ColorWeb/). A copy of these demonstrations are included in the CDROM version of these notes. They demonstrate 1) trichromacy 2) reflectance and 3) metamerism.

Color measurement and evaluation

Instrumentation for tristimulus-based color measurement comes in two forms: instruments that sample spectra and computes tristimulus values based on the samples, and instruments that use filters to approximate the color matching functions. The first usually have "spectro-" in their name, and are either spectroradiometers (can measure lights directly and absolutely) or spectrophotometers (measure relative reflectance, and often contain their own light source). The second form are colorimeters, and are often specifically designed for monitors, reflection prints, or transmission prints. In general, colorimeters are cheaper, faster and less accurate than spectrally-based systems.

All color instrumentation comes with a specification that tells its precision and the range of brightnesses it can accurately measure. Pay attention to this information. There is no point in agonizing over a 1% luminance variation, for example, if the instrument is only accurate to 2%. Similarly, measuring something that is too bright with a monitor colorimeter may simply generate bad data, not an error message. It may also damage the instrument.

All instrumentation needs routine calibration and evaluation to be sure it is working to spec. As well as following the manufacturers recommendations, try to create some standard measurements, and make them routinely to see if instrument has drifted. Similarly, for any task, try to find some "reality check" measurements to ensure you don't waste time gathering garbage. For example, when measuring a monitor, the sum of the tristimulus values for red, green and blue should equal the tristimulus values for white.

Most color measurement equipment can be connected to a computer via a serial port. The vendors often offer both an SDK and some end-user software for color measurement. In general, this software is expensive and bad—these folks understand color measurement, not software. Similarly, be prepared for some quirks in the serial port driver, as the "standards" in this area seem to be rather loosely interpreted, a problem that is not unique to color measurement equipment. It is most convenient if the interface can be driven by typing text codes down the serial port line. This is generally robust, and can be easily interfaced to any system, or even driven by hand from a terminal emulator.

David Brainard and Denis Pelli maintain the PsychToolBox website (www.psychtoolbox.org). The PsychToolBox is a large set of Matlab routines, designed primarily for vision research. However, they include a number of useful routines for general colorimetry, as well as a good page on instrumentation and measurement methods (www.psychtoolbox.org/measure.html).

Densitometry is another form of color measurement, but it is not based on CIE colorimetery. Instead, it measures the amount of light transmitted through or reflected from a material and encodes it logarithmically. Density is defined as $\log_{10}(1/T)$ or $\log_{10}(1/R)$, where T or R lie in the range (0,1). Therefore, a transmittance of 1 is a density of 0, a transmittance of 0.001 (0.1%) is a density of 3.0, and the density of 0 is undefined. Because densities are logarithms, combining two filters sums their density values. Density measurements, either transmissive of reflective, are commonly used for control and evaluation in both print and photography.

A visual densitometer includes a filter that approximates the CIE luminous response curve, making the measurements a function of relative luminance. When measuring colored materials, this creates values that more accurately represent the perceptual lightness/darkness of the sample.

Color densitometers measure density through narrow-band color filters. These are designed to selectively measure colorants in photographic or print media. They have nothing to do with colorimetric values, and cannot be used to compute tristimulus values.

Color Appearance

The models used in colorimetery are remarkably simple. Color is a three-dimensional, linear system that can even be encoded in instruments. However, the truth about the perception of color is much more complex. Color matching as described above only works for individual colors viewed under simple, controlled circumstances. It's definition is based on the simple perception of colored light. In the real world, we see complex arrangements of colors. These colors interact in ways that defy the simple model presented by colorimetery.

In the slides is a classic photograph taken by John McCann. In it, two standard color charts are seen, one in the sun and one in the shade. The difference in lighting is such that the measured intensity of the black patch in the sun is exactly the same as the white patch in the shade. That is, two instruments would assign them the same color, but they clearly look black and white in context. A similar situation exists with the gray step wedge, which lies across the shadow boundary. In spite of the fact that the intensity of the light reflected from the patches actually increases in the sunlight, it appears to show a monotonically decreasing set of gray steps.

Why the appearance of colors in the real world is so different than the simple models defined by the retina is a combination of perceptual and cognitive effects. The perceptual effects are created by the encoding and processing in the brain of the original retinal signals, which is more complex than the simple RGB encoding described so far. Cognitive effects are based on our knowledge of how objects and lights behave in the world.

If you ask people to describe a color, they do not normally speak in terms of its red, green and blue components. They will typically name a hue, such as red, purple, orange or pink. They may describe it as being light or dark. Or, they may called it vivid, colorful, muted, or grayish. This perceptual organization of color is used to define most "intuitive" color spaces.

One of the best known perceptually organized color space is the Munsell space, which is roughly spherical in shape. Colors are organized with a vertical lightness axis (called value), hue arranged around the circumference, and chroma (or saturation) defined radially. Each step in the color space is of equal perceptual distance. That is, the difference between two colors is uniform throughout the space. This is not true of the CIE tristimulus space. The original Munsell definition was created with hand-painted color chips, and you can still buy *The Munsell Book of Color* from the Munsell Color division of Gretag-Macbeth (www.munsell.com). The Munsell chip set has also been defined with respect to the CIE tristimulus space.

There are two uniform color spaces defined by the CIE for the measurement of color differences. Called CIELAB and CIELUV, they are non-linear transformations of the CIE tristimulus values. Like the Munsell space, they are perceptually uniform. They have been defined such that a unit step in the space is considered a "just noticeable difference" or JND. Each space has a lightness axis (L*) and two opponent color axis (CIELAB = L*a*b* and CIELUV = L*u*v*). The equations that define the spaces include the tristimulus values for a reference white. This is necessary to create a single lightness axis.

This perceptual organization of color reflects the first level of processing applied to the cone response, called the *opponent color model*. In this model, the RGB values from the retina are encoded as lightness plus a red-green and a blue-yellow color difference axis. This organization of color can be seen in the afterimages produced by staring at brightly colored patches. Stare at a red color, then look at a white surface—you will see a greenish afterimage caused by saturating the red-green opponent channel. You can create similar effects with blue-yellow patches, and with light-dark ones.

The opponent color model also well describes the common forms of color vision deficiencies. While these are usually caused because one type of cone in the retina is either missing or weak, they are most easily described in opponent terms. The most common problems are anomalies in the red-green opponent channel, where either the ability to see red or to see green is impaired. This type of deficiency appears in approximately 10% of men. A much smaller percentage (1-2%) are weaknesses in the blue-yellow channel, with very few people actually "color blind," or unable to see any hues at all. While most color vision problems are genetic, they can also appear as a side-effect of medication or illness.

Opponent color theory was first proposed by Herring, in 1878, but was not accepted until Dorthea Jameson and Leo Hurvich performed controlled experiments in 1955 that verified the theory. It is now accepted as the first-level processing of the color vision signal after the cone response. Leo Hurvich has written a textbook, *Color Vision*, that is based on this theory. It is also the best technical reference I have found for the modalities of color vision deficiencies.

Color appearance phenomenon are caused by the relationship between colors, by adaptation and by cognitive effects based on our knowledge about lights and objects in the real world. *Simultaneous contrast* describes the influence of surrounding colors on the perception of a color. There are several simultaneous contrast examples in the slides. The simplest model for simultaneous contrast is that the afterimage of the surrounding color is added into the perception of the surrounded color. Therefore, a gray patch on a dark background looks lighter (add white) than a gray patch on a white background (add black). However, the real answer is more complex. In the slides are a set of colored patches on a background of alternating colored bars. Which bar the patch appears on changes its appearance, even though the actual surround is identical. Clearly the cognitive effect, in which we see a patch as being "on" a particular colored bar, is contributing in this example.

Even a thin outline of a different color can make a big difference in the appearance of a colored region. The slide labeled the "Bezold Effect" demonstrates this. The only difference between the two patterns is the color of the outlines; white in one case and black in the other. But, the overall difference in appearance is quite dramatic. This effect was named for a 19^{th} century rug designer, who was able to reuse his designs by this simple alteration.

Size, or spatial frequency, has a strong impact on the perception of a color. The higher the spatial frequency, the less saturated the color. This is why the paint on the sample chip looks less vivid than it does on the wall. In the slide, the two colors really are identical.

Human vision is very adaptable. We are capable of seeing in both very dim and very bright light. When we move from bright to dim lighting, or vice versa, we can feel our visual system adapt. Think of this as a gain control on the visual system. For dim lighting, we turn up the sensitivity. For bright lights, we need to damp it. These phenomena are called *dark adaptation* and *light adaptation*, respectively. In many cases, it is valid to assume that color appearance is relative to the current black/white levels, and is not effected by the absolute luminance. However, increasing luminance can produces a measurable increase in colorfulness (Hunt Effect) and contrast (Stevens Effect).

Chromatic adaptation describes the visual system's ability to adapt to the color of the light illuminating the scene. Most color is created by shining light off of objects. While the reflected spectrum can be measured with colorimetric instruments, changing the light will change the measured color, sometimes dramatically. But, as we view the world, we don't generally perceive objects changing color as the light shifts. Think of it as an automatic white-balancing function for the visual system. That is, the gain controls for the three cones are adjusted separately. Modeling chromatic adaptation is very important for the accurate reproduction of images, which will be discussed in the color media and management part of this tutorial.

The cone response curves physically adapt to the color of the light source. This idea was first proposed by von Kries in 1902. von Kries did not present a mathematical model, but simply proposed the concept that the cones could independently adapt. Modern interpretations of von Kries are expressed in terms of a scalar value, k, that is the inverse of the maximum response for each cone, typically the response for white.

A color appearance model converts from measured values (such as the tristimulus values of the sample and its surrounding color, the overall illumination level, etc.) to correlates of perceptual attributes of color, including hue, chroma, lightness, brightness, etc. All color appearance models include a way to compensate for adaptation to different lighting conditions. Color appearance models are important for digital imaging applications, and are under active development. The CIE has codified a model, called CIECAM97s. Other models include CIELAB (minimal), RLAB and LLAB (improved CIELAB), Hunt, Nayatani, Guth and ATG. Mark Fairchild's book is an excellent reference on this topic.

Lightness scales

Lightness, brightness, intensity, luminance, L*—all are ways of arranging a set of colors along a scale from dark to light. As the control of this mapping is key to many color imaging applications, it is important to understand what these quantities mean, and how they relate to the perception of color.

Lightness and brightness are qualitative terms, with no specific measurement associated with them. Color scientists when speaking formally treat brightness as an absolute term, associated with emissive sources such as lights. Lightness is a relative term, typically associated with reflective surfaces, but can also mean the brightness with respect to a defined white for an emissive display. Informally, they are often used interchangeably.

Intensity is a measured quantity whose units are light power. A typical spectral distribution function plots the intensity at each wavelength. The total intensity of the spectrum is the integral of the function, or the sum of all the wavelength powers. Luminance is intensity measured relative to the eye's spectral response. The CIE color matching function corresponding to Y is effectively the luminous response curve, which peaks in the yellow-green part of the spectrum and tapers to zero at both ends. Take any spectrum, multiply it by this curve, and integrate to get the luminance, which is a measure of the perceived intensity. To better understand the difference between intensity and luminance, consider two lights of equal intensity but of different colors, one green and one blue. The green light will have a higher luminance than the blue one; it will appear brighter for the same intensity because the eye is more sensitive to its wavelengths. Double the intensity of both lights (that is, uniformly scale the spectra) and the luminance will also double. For any given spectrum, there is a scalar relationship between intensity and luminance.

Luminance defines a power scale for perceived color. It can be defined in absolute units, typically candelas/meter^2 But for many applications, only relative luminance is importance, so the luminance measurements will be normalized with respect to some maximum value, typically the value for white.

Perceptually uniform lightness scales

Plotting equal steps of luminance does not create a perceptually uniform lightness scale, where each step is equally different from the other. One good way to define such a scale is to ask an observer to find the gray color half way between two samples, one black and one white. Then split the difference between the gray and the black, similarly the gray and the white. Repeat, until the patches are just noticeably different from each other. The resulting scale will be perceptually uniform.

This process is easy to implement, and people are quite consistent about which values they pick. Another consistency is that they normally pick around 100 patches for the completed scale. This is how the Munsell value scale was defined, from a set of painted samples. The quantity L*, which is the lightness axis of both the CIELAB and CIELUV color spaces, is a function of relative luminance that closely matches the Munsell Value scale. The formula for L* is:

$$L^* = 116(Y/Y_w)^{1/3} - 16$$

$$L^* = 903.3, \text{ for } Y/Y_w \leq 0.008856$$

Where Y is the measured luminance, and Y_w is the luminance of white

What this formula says is that L* is computed as relative luminance (0..1) raised to the 1/3 power, scaled so that a maximum L* value is 100. To avoid the steep slope of the power function as it converges to infinity at zero, L* is approximated with a straight line for small values. The power function is scaled and offset to match the slope where they join. While this seems a bit of insignificant mathematical convenience, it has a substantial effect on the shape of the resulting function. While the formula includes an exponent of ⅓, the resulting function has an effective exponent closer to 1/2.43.

Most pixel encodings of images are non-linear in a similar manner. While the offset, linearity and exponent all vary, the result is that pixel values encode perceptual steps, rather than luminance (intensity) steps. Similarly, working with pixel values (as in an image editing program) that are directly displayed on a monitor with a well-defined gamma curve is equivalent to working in a similarly encoded non-linear space.

Non-linear spaces can be very confusing, so let's work through an example. If you have a step wedge from black to white representing reflectance values such as 10%, 20%…100% with respect to the black (i.e., define black as 0%), then the patch that appears halfway between black and white has a reflectance of roughly 18%. If you give this patch an encoding of 0.5 (out of 1.0), then you need more dark patches and fewer light patches to create a perceptually uniform scale. Or, you can think in terms of a light with a variable step intensity setting—it's roughly equivalent. The precise perceptual mapping depends on a number of variables like the size and brightness of the stimulus and the surround, but the non-linearity is similar.

Now, display a black, white and gray patch on a monitor whose response is a non-linear gamma curve around 2.2-2.4, and adjust the gray patch until it appears half way between black and white. The pixel value you select will be around 0.5, depending on the actual value of gamma. This is because the monitor's response curve also assigns more pixel counts to dark colors than light ones. So, working on an "un-gamma corrected" monitor is the same is as working in a non-linear, (roughly) perceptually uniform color space. There are some diagrams in the slides that show this concept schematically.

Lightness and objects

The lightness of objects is described in terms of their reflectance or transmittance, which indicates the amount of incident light reflected from or transmitted through the object. These quantities are most precisely described as spectral distributions, but it can be useful to describe the total reflectance or transmittance as a percentage, especially when comparing object colors with similar spectral characteristics. For example, a set of gray filters of varying degrees of darkness could be described as having transmittance values of 25%, 50% and 75%. But, this is only precise if the transmittance spectra are scaled multiples of each other. In some applications, the luminous response curve is applied to arbitrary reflectance spectra and normalized so that a maximum luminance value is 100 (for 100%). This a somewhat questionable practice, and only corresponds to luminance if the object is illuminated with a light source that has equal energy at all wavelengths.

Density measurements of materials create a logarithmic lightness scale based on the relative reflectance or transmittance. Such a scale is perceptually more uniform than a linear scale of the same value.

In color imaging applications, the mapping from pixel values to perceived lightness is an important aspect of the image appearance. Controlling this through the image reproduction process is called *tone reproduction*, and is one of the key components that affects image quality.

Color Reproduction and Management

Color reproduction focuses on the reproduction of colored images. It has been defined by the different technologies used in the traditional color reproduction industries: television, printing and photography. The best reference for this topic I know is Hunt's *Reproduction of Colour.* First published in 1957, the fifth edition was published in 1996. Dr. Hunt is still active in the field, working with the *Colour & Imaging Institute* at the University of Derby in England. Gary Field has two newer books on the topic, *Color and Its Reproduction, 2nd Edition* and *Principles of Color Reproduction*, an update of John Yule's classic book on printing, which has long been out of print.

Traditionally, each color reproduction industry had its own separate technologies and skills. With the coming of digital imaging systems, these industries are being integrated, both with computing and with each other. It is not unusual to find a graphic arts business equipped with digital scanners, printers, film recorders, video cameras and video editing stations. The representation and management of color images in digital form was hinted at in the late 70's, demonstrated in the early to late '80s, and is now an established commercial domain. *Color management systems* are now part of every high-end color reproduction application, and are becoming more common for simple desktop publishing users. Giorgianni and Madden, both from Kodak, have written a book called *Digital Color Management* that is an excellent reference for this topic from the system and implementer's viewpoint. From a user's viewpoint, I recommend Phil Green's book, *Digital Color* (highly scan/print oriented) or *Adobe Photoshop 6.0 for Photographers* by Martin Evening.

The color reproduction industries, photography, printing and television, were traditionally focused on the reproduction of natural scenes. Such reproductions are inherently approximations, as no technology can reproduce the full range of colors visible in nature. Each industry tuned its process to provide the best approximation of the real world on its medium. Especially in the graphic arts, any part of the color reproduction pipeline could be adjusted to achieve the desired result.

Now, there is more of a focus on cross-media reproduction and the integration of purely digital content as well as images of nature. This means that color reproduction is becoming less about specific technologies and more about integrating all of them. This has created an increased dependency on standards and process control, to isolate the aesthetic and image-dependent variations from process variation.

Color reproduction

The process of color reproduction involves capturing the light reflected from or emitted by an original, then reproducing it for viewing. The goal is to make the reproduction look as much like the original as possible. That is, viewing the reproduction should be as similar as possible to viewing the original scene, within the limitations of the medium. All media can only reproduce a fraction of the human color vision experience, so all reproductions are approximations. The definition of the ideal reproduction ultimately becomes one of preference and context, though there is substantial technical and craft information about creating good reproductions. As well as the books mentioned in the introduction to this section, there are the publications of the Society for Imaging Sciences & Technology (IS&T) and those of the Graphic Arts Technical Foundation (GATF).

From our study of color matching, we know we can encode the colors reflected from a scene as three numbers that represent roughly the red, green and blue components of the colors seen. *Image capture,* therefore, is the process of capturing and encoding these values for every point across the image. These are the familiar image pixels used in digital representations of images. The separate red, green and blue images are sometimes called *color separations*. These separations are reproduced by mixing primary colors for the output medium. For light-producing media such as monitors or digital projectors, the primaries are red, green and blue light. This is called *additive reproduction*. For subtractive media such as film and print, the process involves subtracting the red, green and blue components from a white light using layers of colored filters, or *subtractive reproduction*. The primaries for the subtractive media are named after the color of the filters: cyan, magenta and yellow.

Good image reproduction maintains the appearance of the original. As in color appearance models, control of *tone reproduction*, which is the reproduction of the overall brightness and contrast (the achromatic axis) is separated from the problem of adjusting hue and saturation. Tone reproduction is simpler to model than color reproduction as it involves only one dimension and can easily be implemented with table lookup in digital systems. Many image reproduction problems can be satisfactorily solved simply by controlling the tone reproduction. In traditional image reproduction industries, careful tone reproduction was augmented with color correction, a step that simply tried to adjust the primary input colors to better match the output media. This was usually achieved by processing each separation independently, as can be done with the RGB (or CMYK) curves in tools like Adobe's Photoshop.

Color management systems provide a more general mechanism for matching colors based on measurements of the *gamut*, or set of all possible colors that can be produced by any output device (printer, monitor, etc.) Generating such a specification is called *characterizing* the device. The characterization data is encoded in a *device profile* that represents the color reproduction characteristic for the device. Modern color management systems use a colorspace based on the CIE standards for this purpose. Most commercial systems use CIELAB, though some also support the use of the standard CIE XYZ tristimulus values. Future systems will hopefully be based on appearance models such as CIECAM97s.

Image capture

To perform image capture, the scene is viewed through red, green and blue filters. If the filters represent color matching functions, the red, green and blue values would be tristimulus values. Typically, however, image capture filters are not color matching functions (or *colorimetric filters*). Such filters are expensive to make and inefficient to use because they absorb a great deal of the incoming light. More often, the filters are tuned to match the expected inputs and outputs. Because they are not true color matching functions, they may mis-code some colors. That is, two colors that look different to a human viewer may encode to the same RGB triple, and vice versa. However, substantial engineering effort has been spent to minimize these effects. Good quality video cameras, for example, contain filters that are close to colorimetric within the expected video display gamut, but not outside. This means that the biggest miscoding errors affect highly saturated colors that are difficult or impossible to display anyway.

For use with color management systems, the captured image must be characterized with respect to some known device-independent color space. This is accomplished either by characterizing the input device, or by displaying the image in some standard manner and characterizing it with respect to the display system.

Input device characterization can only be general if the device uses colorimetric filters. Otherwise, the characterization depends on the input spectra used in the measurements.

For example, a scanner could be characterized by scanning a set of colored patches with known CIELAB values and making a correspondence between the resulting RGB pixels and the CIELAB values. Other values could be interpolated from this set, and the inverse transformation generated by inverting the input table. However, patches of a different material (paint vs. a photographic print, for example) that have the same CIELAB value might produce different RGB values in the scanner. Commercially, most scanned material is photographic prints or other printed materials of a standard form so that there are a limited number of characterizations needed in practice.

Camera-style image capture devices (video and still digital cameras) are typically used to capture arbitrary spectra from nature. It is more common, therefore, characterize these images with respect to a display. Video color specifications define the camera filters with respect to a standard display in a way that automatically provides this characterization. Images captured from most digital still cameras, however, require some adjustment after capture. It makes most sense for these systems to characterize the image with respect to the user's display.

Additive color systems

Additive color systems reproduce the red, green and blue parts of the image by adding together red, green and blue lights. Such systems include monitors, liquid crystal displays (LCD) and digital projectors. The traditional additive color reproduction industry is television, though scanned graphics displayed on a monitor may be more familiar to this audience. Charles Poynton's book on *Digital Video* is a good reference for video and television technology. Lindsay Macdonald and Anthony Lowe have edited a good general reference on *Display Systems* for those who want more detail about specific display technologies. The Society for Information Display's (www.sid.org) conferences and publications are the standard venue for research in display technology.

To achieve a good additive reproduction system we need to match the color and the relative brightness of the input and output primaries. Let's talk first about intensity, or brightness. The *intensity transfer function* (ITF) is the curve that maps input values (pixels or voltages) to intensity (viewed light). A monitor's ITF is a non-linear, power function called a *gamma function*, which is defined primarily by the physics of the electron beam that excites the phosphors. Most modern monitors are manufactured such that they have identical gamma functions. The shape of the function can be modified, however, simply by changing the contrast and brightness adjustments on the display.

In the slides is a diagram that describes the ITF's for the television pipeline. The linear intensity values reflected from the original scene are non-linearly encoded in the camera. This both optimizes the use of pixel values by companding or compressing the intensities, and creates values that can be fed directly into the television electronics. The original design was to make this encoding is exactly the inverse of the gamma function of the receiving monitor. However, studies by Bartleson and Breneman in the 70's demonstrated that a slightly non-linear reproduction was perceptually more accurate because of the normally dim viewing environment for television. Therefore, while a typical monitor has a gamma function of 2.5, the camera encodes to the inverse of 2.2. Similarly, most recommendations for "gamma correction" use a value of 2.2 as well.

Recently, the issue of 2.2 vs. 2.5 has taken a new twist. Rather than attributing this difference to perception and/or viewing environment, some color scientists are claiming that the "true" value of the monitor gamma is 2.2, once you account for inner reflections and other variables that affect how the light is emitted from the front of the display. What this all proves is that it is very difficult to decide how and what to measure, even for such a simple question as the display transfer function. In practice, assuming that a properly adjusted monitor's gamma is (effectively) 2.2 is sufficient for most applications. Note the "properly adjusted." Most "gamma" variations are actually caused by curves that are not gamma functions because the brightness and contrast settings have drifted over time or by user intervention.

For monitors, the color of its primaries is defined by the phosphors used. Most modern monitors have a similar set of primary colors. Liquid crystal displays, however use colored filters and a backlight to define their pixels. Standards for these are still evolving. Specifically, there is a tension between the desire for bright saturated colors (bright, high-powered backlight, intensely colored filters) and battery life (dimmer, low-powered backlight and thinner, less saturated filters). Recent measurements of desktop flat panel displays show that the red and green primaries are approximately the same as monitors, but that the blue value is less saturated. Laptop displays, however, are substantially desaturated with respect to the standard monitor phosphors. The ITF for a liquid crystal display is not inherently a gamma curve, though most manufacturers apply one artificially for compatibility with monitors and video.

Digital projectors are another additive color device that are becoming common. A projector contains an imaging element such as a small LCD or an array of micro-mirrors (DMD) that modulates the light from a high-intensity light bulb. Most LCD projectors and the larger DMD projectors contain 3 imaging elements and a dichroic mirror that splits the white light from the bulb into its red, green and blue components. These are recombined and displayed simultaneously. The smallest projectors, however, use a single imaging element and a color wheel of filters, so the separations are displayed sequentially. The DMD projectors based on Digital Light Processing (DLP) from Texas Instruments can weigh as little as 3 lbs and still produce a bright, crisp, image. Their colorwheel includes a clear filter that is used to add white to the brighter colors. This increases the dynamic range of the system, but makes it more complex to characterize. The native ITF for a projector is defined by the imaging element. DMD elements actually pulse encode the grayscale, making them naturally linear. However, like flat panel displays, most projectors contain image processing hardware that induces a gamma curve compatible with monitors and video.

There is are diagrams in the slides that plot the typical monitor gamut (yellow solid triangle) with several different LCD gamuts and a couple of digital projector gamuts. Traditionally RGB reproduction systems assume a common RGB because most displays were monitors. Now, this is no longer true. However, it is straightforward to transform from one RGB system to another by inserting a matrix multiplication in the output pipeline. This can be implemented in special purpose hardware, as in video systems. Modern graphics cards, such as the Nvidia G-Force series (www.nvidia.com), provide a programmable function that can be applied to this problem.

Characterization of additive systems is fairly easy because there is a linear transformation between RGB intensity values and CIE tristimulus values. This transformation can be defined by measuring the red, green and blue primaries. For most display technologies, this creates a 3x3 matrix that transforms from RGB intensity to XYZ. For projection displays, it is necessary to include a translate component to compensate for the high black level of these devices. This results in a 4x4 homogeneous transformation matrix. Note that the transform must be applied after the ITF is applied, not to the pixel values directly. The DLP projectors with the clear filter in the colorwheel, however, are not so simply characterized.

Subtractive color systems

Subtractive color systems filter the red, green and blue components of the image from white light. To do this, they use colored filters that in theory modulate only the red, green and blue components of the spectrum. The filter that passes green and blue but modulates red appears cyan. Similarly the green-modulating filter appears a purplish-red called magenta, and the blue-modulating filter is yellow. Therefore, the primaries of a subtractive reproduction system are said to be cyan, magenta and yellow. In printing, black ink is added as well, to improve the contrast.

The advantage of subtractive color systems is that they use a single, white light source instead of three colored ones. Furthermore, for reflection prints, this light is simply the light in the room. While flat-panel displays are getting thinner and lighter, it is hard to imagine they will ever be as light, convenient and inexpensive as a sheet of paper.

In a reflection print, the colored ink or film forms layers on a white background. Light passes through the layers and is reflected back to the viewer. Compared to light emitted from a display, or even light passing through a colored transparency (like a 35mm slide), prints are low-contrast and desaturated. This is caused by scattering as the light passes through the colored layers twice. Reduced scattering is why colored prints look more vivid on smooth, coated papers than on matte ones.

The cyan, magenta and yellow filters created with printing inks are far from the ideal filters described above. That is, the cyan filter, which should only modulate the red part of the spectrum, also affects the green part. Similarly, magenta modulates both red and blue as well as green. This means that there is interaction, or *cross-talk* between the filters. They do not function as independent primaries in the same way that RGB lights do. The result is that it is difficult to predict what color a particular combination of inks can produce. Whereas RGB systems are linear, CMY systems are not.

In printing, black ink is often added to the cyan, magenta and yellow separations to improve contrast and to reduce the amount of ink spread on the paper. In theory, black and gray colors can be created by combining equal amounts of cyan, magenta and yellow. In practice, such a combination will often produce brown or olive green. *Gray balancing* is the process of determining the correct mixture of primaries to create a gray color. The darkest color produced by mixing the three primaries, black, can be made darker by printing black ink on top of the three primaries. It is also possible to replace some of the colored ink with black, though once again, there is not a simple, linear relationship that predicts how much black ink is needed to replace a specific blend of CMY. The process of substituting black for the primaries in the darker colors is called *gray component re-*

placement, or GCR.

Another topic of significance for printing is *halftoning*, or the process of converting intensity into dot patterns on the paper. Most printing systems are binary—ink either is printed or not. To get the appearance of gray, the ink is printed as high-resolution patterns called *halftone patterns*. Originally made for mechanical printers by using a special filter and films, these patterns are now digitally produced on ultra-high resolution film recorders. The design and implementation of the halftone pattern has an important effect on the appearance of the printed image. Digital printers often use other patterning techniques such as *dithering*, and may have the ability to produce some grayscale on a dot-by-dot bases. For example, an inkjet printer may create each spot with multiple drops. The more drops used, the darker the color.

There is no simple, linear transformation that maps subtractive devices to tristimulus space. Combining filters multiplies their effect rather than adding them. Most subtractive devices are characterized by measuring a large number of color patches then interpolating between them. Anything that changes the output appearance, such as changing the halftoning algorithm in printing, will change the characterization.

Color Management Systems

There are a pair of diagrams at the start of this section (courtesy of Michael Bourgoin of Adobe Systems) that succinctly describe the digital color management problem. Given a set of digital input and output devices such as printers, scanners, monitors and film recorders, how do we convert colors between them without creating a customized transform between each pair of devices? The answer is to use a single, device-independent color space, such as those defined by the CIE, as the common interchange format. For each digital device, there is a profile that maps between its native color space and the interchange space, or Profile Connection Space (PCS). Similarly, images must include implicit or explicit profiles that describe their colors with respect to the PCS.

Many companies produce color management systems, including Kodak, Agfa, Lineotronic and EFI, just to name a few. Apple's ColorSync is built into the Apple operating system in such a way that it can import color management engines from a variety of vendors. To support the exchange of profile information across systems, the International Color Consortium (ICC) was formed. This industry consortium has defined a standard profile format that is now commonly used by all commercial systems (www.color.org).

On a Windows machine, the profiles are .icm files. These typically installed with the device driver for a new printer, monitor, etc. These default profiles vary in how well they approximate the actual device, and of course, the system depends on the user to correctly identify the system components. For displays, the contrast and brightness settings must be correct to get a good match. The setup procedures in Photoshop and similar program are designed to help maintain this. Serious practitioners must either invest in profile generation instrumentation and software, or pay to have custom profiles created.

The role of the color management system is to manage the profiles and to transform images from one device to another based on the profile information. Because there is no single "correct" way to make these transformations unless the gamuts match perfectly, different systems will give different results.

Matching image colors based on the CIE color matching metrics is not really sufficient. Image appearance is complex, and while using CIELAB is better than trying to match tristimulus values, it still falls short of being a true image-appearance space. For example, take an image viewed on a monitor that is set with white at a color temperature of D9300 (a very bluish-white that is the default for many computer monitors). Now, reproduce it such that it looks good viewed in normal room lighting (anything from D5000 or warmer). This requires transforming the image white point in a way that mimics chromatic adaptation. The type white point transformation encoded in CIELAB, however, causes unexpected hue shifts. Future systems may include the use of improved color appearance models such as CIECAM 97. These models, however, are more complex to implement than CIELAB, so industry acceptance may be slow.

Gamuts and gamut mapping

The gamut of a device creates a volume in the PCS. Only colors that fall within this volume can be reproduced on that particular device. The slides show several gamut representations. The scatterplots represent the gamuts as 3D plots, which is the most accurate way to think about them. Shown are a gamut for an additive system (a monitor), which is a regular shape similar to a cube but with parallelograms for faces. This graphically demonstrates that the transformation from the RGB color cube to CIEXYZ is a linear one. The Cromalin print gamut, however, is distinctly non-linear. While there is some hint of the original cube, each face is a different size and some of the edges are distinctly curved. The gamut is actually concave as well as bent. The gamut for the DLP projector shows a basic additive structure plus an extrusion at the white point caused by selectively adding white. As a result, it is linear in some regions and not in others.

Because the data used to generate the plots was absolute measurement data, the scales are quite different. However, in color management applications, this data would be converted to CIELAB, which normalizes it to a common white (but not necessarily black) value. Another figure in the slides shows gamuts projected onto the CIE chromaticity diagram. Here, linear systems project as triangles whose vertices are the chromaticity coordinates of the red, green and blue primaries. The print gamut, in contrast, is quite irregularly shaped.

To map from one device to another, *gamut mapping* is applied as transformations in this common color space. Current practice is to project colors towards the center of the gamut in a way that reduces saturation, and, to a lesser extent, brightness, while maintaining hue. However, if no in-gamut color is really close, than the reproduction will look substantially different no matter how clever the gamut mapping algorithm. Only redesigning the image can really create a good reproduction.

User's view of color management

The color management system implementer thinks in terms of device profiles, profile connection spaces such as CIELAB, and transformations between color gamuts. Most digital color users, however, want to think in terms of a common "working space" that is usually some form of RGB color space. Their images or other digital colors are created and manipulated in this space. Colors and images are imported into this space. Displays and other output devices are used to view this space. The transformations important to the user are those that convert colors and images in and out of this space, and include not only the device profiles, but any settings that affect the transformations using these profiles.

The choice of working space depends on the application, both in terms of constraints on the RGB space (size, linearity), and whether the images and colors created in this space need to be exchanged, or merely displayed and printed.

All color management relies on measured data stored in profiles, either created by the user, or supplied by the manufacturer. The more accurate the profile, the more effective the color management. Few users have the equipment or expertise to profile all of their digital devices, so they must focus instead on maintaining their equipment such that the default profiles remain reasonably accurate.

Picking an RGB working space

An RGB working space is defined by the color and magnitude of its primaries, typically expressed as the chromaticity of its primaries plus the white point and magnitude of white. Its pixel values can either be linear with respect to intensity (or luminance), or non-linear. While in theory both rendering and image processing algorithms should always be performed in linear intensity spaces, the vast majority of users work in a non-linear space defined by their display's characteristic transfer function. While linear spaces are physically correct for some computations, these non-linear spaces are more perceptually uniform for direct manipulation of colors and images.

The ideal primaries for the working space also depend on the application. It is most convenient to make these match the user's display, as this makes the entire working space visible without gamut mapping. However, there are colors in print and film that fall outside a typical monitor gamut (golden yellows and foliage greens most noticeably). Therefore, in the graphic arts, where the work flow often progresses from scanned photographs to prints, it is not uncommon to choose a larger RGB space than that of the monitor to avoid compressing these colors into the monitor's gamut. The cost is that the monitor, or any other common output device, can no longer accurately preview all colors in the working space. Another alternative in the graphic arts is to effectively make CMYK the working space, but that isn't as generally convenient as using an RGB space.

All colors and images created in the working space are defined with respect to the working space. If they are only going to be "exported" as prints or by conversion to some standard format, then by far the most convenient space to use is the one that matches the user's display. If you need a linear space for computational reasons, derive it from the display's RGB space so that only the intensity transfer function varies between them.

Standard RGB Spaces (sRGB)

While video formats are very explicit about the RGB colorspace underlying them, few still-image specifications are so rigorous. One effort to fix this problem has been to establish standards that include RGB specification information. This has had little practical effect yet. Such information is mandatory in PNG, but optional in both TIFF and JPEG 2000 (and missing entirely from the original JPEG file format).

Programs designed for the graphic arts (such as Quark Express or Adobe Photoshop) do routinely create and read file formats that include images with embedded profile information. One of the great complexities in Photoshop color management comes from the fact that that it does allow such images, which can either be converted to the user's working space (which may or may not be their display space) or preserved in their original forms and converted dynamically for editing only. This is needed because conversions may cause lost information. This degree of generality, however, is not always needed and almost always causes confusion. Adobe has tried several times to make this simpler and easier to understand for its Photoshop users, but it is still a problem.

The sRGB colorspace has been designed to address the problem of a standard RGB colorspace for desktop publishing and the web (www.sRGB.org). It is equivalent to a monitor set with gamma = 2.2 and a white point of 6500. This is, not coincidently, similar to the default settings for a modern monitor on a PC system once the white point is set to 6500. This makes it possible to use the standard sRGB space not only as an export standard, but as the user's working space, simplifying matters even further. Be aware, however, that as monitors age (over a couple of years of constant use) both their brightness and color temperature will drift. Therefore, you must find a way to verify the actual gamma and white point, either via measurements or some visible comparisons, before defining your monitor as sRGB compatible.

Macintosh users, however, commonly work in an environment where the display system modifies the effective monitor gamma to 1.8, which is why lists of monitor specifications for color management often include a "gamma 1.8" as well as a "gamma 2.2" default option. The Color Synch monitor calibrator in the Mac OS makes it possible to set gamma to 2.2 on a Mac, but it is contrary to standard Mac practice and may make many displayed colors and images appear too dark. On a Mac, it would be better simply to export in sRGB rather than use it as working space. As only the intensity transfer function is different, this would be a simple conversion.

Flat panel displays, especially laptop displays, are not sRGB compatible because they have different primary colors. However, it is easy to transform these display colors to sRGB because their gamuts fall inside the sRGB gamut.

The sRGB colorspace is similar to the digital video color space (ITU-R BT.709) and uses the same primaries, though there are some subtle differences in the specification of the non-linear transfer curves. There has been some flaming in the graphic arts community about sRGB, primarily for the reason described in the previous section—it is a monitor space, which will clip some print and film colors.

The EXIF image format from JEIDA (Japan Electronic Industry Development Association) combines JPEG encoding with an sRGB colorspace specification for digital photography.

Establishing Accurate Color Transformations

To use color management effectively, you need a well-defined working space and well-controlled transformations into and out of this space. For many users, the simplest solution is to make the display space and the working space the same. However, the display must be characterized, and its characterization maintained so that the working space is well defined. Similarly, all target output devices (printers, film recorders, etc) must be characterized, and any imported colors (images, video, etc) must be accurately transformed into the working space.

These color space transformations are often defined in three steps: calibration, profiling and color mapping. Calibration and profiling combine to create a device-independent specification for the device, its profile. Calibration generally means making sure the hardware is working within spec, and profiling means taking the measurements to map from the device coordinates to the profile connection space (PCS). However, calibration can include setup steps that not only eliminate anomalies (such as ¼ of the monitor pixels mapping to black because the brightness is too low due to aging), but are used to establish a standard form (such as setting the gamma curve for a monitor to 2.2).

In an ideal color management scenario, the user will accurately measure and profile any digital device interest—monitors, printers, scanners, etc., and measure their systems regularly to ensure quality. In practice, this happens only in color critical imaging industries such as special effects or graphic arts production houses. And, few outside these domains have the equipment or expertise to create accurate profiles.

To some extent, color management can be applied "incrementally," with results proportional to the amount of effort applied. It is relatively easy to approximate a display profile using a combination of manufacturer supplied data and visual estimations for gamma. The Adobe Gamma Utility, which is bundled with many of their products, provides a reasonable way to establish gamma visually. On a PC, it will actually adjust internal tables in the display controller to establish the desired gamma curve. Packages that include a display measurement tool and software for profiling can be had for under $500.

Digital desktop printers come with surprisingly good default profiles and algorithms for transforming from a display profile to a print. To get good results, you simply need a reasonable display profile for your system and the manufacturer's data for the printer. Then, be sure color management is turned on in the print driver. The major pitfall in printer calibration comes from the multitude of settings, paper types, etc. all of which affect the image color. Manufacturers are unfortunately not specific about which settings will override the profile, or which are assumed. Ultimately, gamut mapping from monitor to print will have the biggest effect on the color.

As described above, many image formats specify their RGB colorspace, but many do not. In general, images created by cameras will be non-linear, and will assume something similar to the "standard" phosphors defined for sRGB. Computer graphics formats, however, may be linear. As described above, any image or individual colors created on a monitor should be treated as non-linearly encoded. The sRGB encoding is a good default for PC's, but Mac's have a different transfer function. Colors defined on a flat panel need will also be non-linear, but have different primaries.

Color Synthesis

In computer graphics, we are often involved in the synthesis of color. Rendering systems synthesize color from the interaction of lights and surfaces. How the color is specified and represented depends on the rendering model and application. Rendering algorithms are part of graphics systems, where color is represented in a file, rendered, displayed, and integrated with colors and images from other applications. Application designers often provide users tools for selecting individual colors for illustrations, slides, web pages, or other designs. Graphic and information visualization designers have rules and systems for selecting colors, sets of colors and their application.

This section of the talk will discuss all of these areas, which will be presented more as a set of topics rather than as a single, well-integrated story. There are three main sections: Color in rendered graphics, color selection and color design.

Color in Rendered Graphics

Colored lights stimulate the eye directly. However, the light we see when we look at an object is the product of the emission spectrum of the light and the reflectance spectrum of the object. In the Color Playground is a tool that multiplies two spectra to show the simplest form of this effect. In nature, however, interactions between lights and objects is more complex. Shiny surfaces have highlights, rough ones appear textured. Particles and layers as small as the wavelength of light produce shimmering interference effects, like a butterfly wing. Light shining through clouds of small particles produce the blue-gray of smoke, the colors in the sky, and the blue in a child's eyes. It is the goal of rendering research to produce effective models of these processes that can be used to make compelling and beautiful images and animations.

Unlike image reproduction, where the focus is on producing images of objects in the world, computer graphics is focused on producing images of objects that occur only as the result of numerical models. That is, there are virtual objects, and virtual lights that must be rendered to produce a visible image. How they are modeled and rendered defines their appearance. One of the most widely referenced books on color in computer graphics is Roy Hall's *Color and Illumination in Computer Graphics Systems*. Unfortunately, this book is out-of-print. However, most books on rendering include a discussion of color. I have used Andrew Glassner's *Principles of Digital Image Synthesis* as my reference for rendering.

Simulation vs. Design

Philosophically, computer graphics rendering can be organized between two extreme positions that I call simulation and design. The goal of simulation is to model reality. The emphasis is on physically correct representations and rendering algorithms, which for color means at minimum a spectral model of light and surface reflectance. The goal of design, on the other hand, is to produce a visual effect. Most design systems have lots of knobs, and the designer may do unreal things to create the desired result. One of my favorite stories demonstrating this involves an early graphics system that could not produce shadows. Yet the designers were making pictures with shadows in them. How did they do it? They added negative lights. Because the emphasis in design systems is interaction and flexibility, they typically use computationally efficient, RGB-based models.

In simulation systems, the result should be good enough to measure. Rendering involves computing the interaction of light with a material surface. Both the surfaces and the lights should be physically accurate, preferably created by measurement. The equations that control the interaction and scattering of the light should also be physically accurate. Radiosity, and more broadly, global illumination models, are required to describe the interaction of light throughout the scene. To define the visual representation of the image, tristimulus theory at a minimum is used, with more advanced perceptual models a topic of current research.

In design systems, what the final image looks like is the goal. Unreal may be good, as in the negative lights example above. Designers don't object to using physically accurate models, as long as they are efficient and flexible. Because performance is a paramount concern, simple lighting and shading models that can be accelerated with special hardware are the most common in design systems. Most of these have their basis in Phong's original shading model, which is extended to color simply by applying it to RGB independently. Paul Strauss created an extension of this that has become part of SGI's Open Inventor (www.sgi.com), which is a widely-used 3D graphics format. The designer can either simply paint surfaces or vertices with RGB colors, or apply simple material and lighting models. Color is specified for the surface, its highlight, and its ambient (shadowed) components. Lights, too, are colored based on an RGB specification. RGB or intensity textures can be applied to surfaces. In his original paper, Strauss stated that this model was created to give designers a useful set of parameters for controlling the appearance of their graphics.

A*rtistic* or *non-photorealistic rendering* applies painterly and illustrative style and techniques to computer graphics. Gooch shading, presented at SIGGRAPH '98, is a variation of the Phong shading model that mimics the type of shading a professional illustrator would use. It does not represent a realistic lighting model, but is optimized to illustrate the shape of the object being shaded. Seminal papers from the University of Washington have demonstrated rendering techniques that produce images that look like drawings made with traditional media such as pen and ink or watercolor. Amy and Bruce Gooch have now written a book, *Non-photorealistic Rendering,* which is listed in the bibliography. There is a bi-annual conference, Non-Photorealistic Animation and Rendering (N-PAR), and much other activity in this area.

The commercial program *Painter* (was Fractal Design, is now Metagraphics at www.metagraphics.com) is a 2D painting package that does an excellent job of mimicking pens, pencils, paints, and the like. It includes the effect of textured paper as well. While the designers of Painter have not published their algorithms, most systems that model ink and paint rely on the Kubulka-Monk model to create their effects. The Kubulka-Monk model describes paint (or any other similar colorant) as a sequence of thin layers. The model describes a way to compute the scattering and absorption in these layers. By varying the thickness of the layers and other parameters, you can get a good model of painted materials.

RGB vs. Spectral Representations

Within rendering systems, the two main ways to represent color are as spectra or as RGB triples. A spectral representation provides the only accurate way to model the interaction of light, direct or reflected, with object colors. An RGB triple can be thought of as a tristimulus representation for the color, or as a view based on color separations. However, while colors can be accurately represented by such triples, their interactions at the spectral level cannot. That is, you can't predicted the result of multiplying spectra by multiplying their tristimulus values. Therefore, RGB representations are inherently inaccurate in rendering systems except for the simplest cases. In general, the more spectral computations needed (to model reflections, for example) the less satisfactory the RGB representation.

So why not always use spectral representations? Because they are expensive. The range of wavelengths that comprise the visible spectrum runs from 370-730 nm. Good color measurement instruments sample this at 2 nm intervals, though many sample at 10 or even 20 nm intervals for efficiency. An equally-spaced sampling of the visible spectrum, therefore, contains from 180 to 18 samples that must be stored for every color in the screen and processed for every rendered pixel. And, for color design and selection, spectra provide a very unintuitive model. After all, human color vision starts with a trichromatic encoding of the spectrum—we do not perceive spectra directly.

It is possible to generate more efficient representations than uniform sampling if we know something about the possible shapes of the curves we want to represent. The surface reflectances of non-fluorescent objects[2] are often very smooth, as is the spectral distribution of daylight or of incandescent light. Therefore it is possible to adapt the sampling to give a more efficient representation. Gary Meyer, Roy Hall and Andrew Glassner have all proposed methods for doing this, and Roy Hall has a paper surveying this topic in the July 1999 issue of *IEEE Computer Graphics and Applications*. The most efficient representations use as few as 4 samples/spectrum. Note, however, that a more concise representation may mean a more expensive reconstruction algorithm when the spectra are actually applied in the rendering.

Another way to create an efficient representation for a set of curves is to define basis functions that can be linearly combined. With this approach, only the scalar multipliers for the basis functions need to be stored for each color. Maloney and Wandell have applied this technique to a variety of surface datasets, and Mark Peercy has applied this approach to computer rendering in his SIGGRAPH '93 article.

To display spectra on a monitor, they must be converted to RGB monitor values. Using trichromatic theory and the CIE color matching functions, it is straightforward to convert from spectra to tristimulus values. Because a digital monitor is also a three dimensional additive color system, there is a linear transformation that converts from CIE tristimulus values to the intensity needed for each of R, G and B. This matrix is derived from the tristimulus values that describe the monitor primaries, as is shown in the slides.

RGB representations evolved by simply applying grayscale shading models to the red, green and blue image components independently. The effect is that any color is described as a linear combination of the three display primaries. If all we want to do is specify individual colors, or additive mixtures of them, this is fine. However, in rendering, we multiply the spectrum of the light by the reflectance of the surface. This is not guaranteed to produce another spectrum that can be represented as a linear combination of the display primaries. Said another way, if the RGB values are the tristimulus values for the spectra, the product of the tristimulus values does not equal the tristimulus values of the product of the spectra. The exception is when one of the spectra is an equal energy white, which

[2] Phosphors, which are used in CRT displays or in the coating of fluorescent lights, contain high-energy spikes that must be carefully sampled.

is often the case for simple graphics systems. A SIGGRAPH ’91 paper by Carlos Borges provides an analysis and bound for the error created by using an RGB model.

Using an RGB representation for color can be very effective. Its failings only really appear as the lighting and shading models become more complex. The difficulty of constructing a simple case where RGB fails can be seen using the demonstrations in the Color Playground from Brown University that are included with the CDROM version of these notes.

Viewing Rendered Graphics

All graphics rendering algorithms must ultimately produce an image that can be displayed on a monitor or some other digital display device. The rendering color space is an unbounded, linear intensity space that must be mapped to the bounded, non-linear color space of a display system. As in image reproduction, the issues include definition of the correct RGB primaries, intensity mapping, out-of-gamut colors, and appearance considerations.

Rendering systems that use spectral representations are specified absolutely. However, those represented as RGB triples typically are not. For graphics interchange, such as the VRML graphics standard, it is critical to complete the calibration. Otherwise, it is impossible to make the rendering look the same across two different display devices. While many monitors have similar RGB primaries, LCD panels and digital projectors do not. Also, the RGB specification in the rendering space is linear with respect to intensity, whereas most display systems are not. Therefore, both the RGB mapping and the intensity transfer function (ITF) must be determined to create a correct rendering.

Conversions between RGB systems can be accomplished with a 3x3 matrix transformation. For maximum accuracy, this should be performed per pixel. However, as most rendering systems interpolate colors as they fill, providing the transform per vertex would probably be an acceptable approximation. Unfortunately, this is more computation than most graphics systems are designed to apply to this problem.

The intensity mapping can be controlled with 1 or 3 lookup tables. Classic 3D graphics systems are stand-alone applications. They can use the lookup tables in the display controller to control the ITF of the displayed pixels. This is sometimes called “gamma correction” because the goal is to “correct” the non-linear gamma curve of the monitor to be the linear response defined by the rendering system. However, a linear ITF is not perceptually uniform, nor is it an efficient encoding of an 8-bit pixel. Therefore, most desktop systems are not “gamma corrected” to linear intensity. Furthermore, different operating systems set different default ITF’s. As a result, colors on a Macintosh are lighter than those on a PC (actually, it’s a matter of contrast) and darker than on an SGI workstation. There is a slide that summarizes the cross-operating system gamma issues. Charles Poynton has also written extensively on this topic.

In a multi-window, multi-application system, it is appropriate that the linearity of the display be controlled by the operating system because it is a shared resource. Therefore, the issue is not whether the lookup tables in the display controller can be changed by the 3D rendering application, for even if they could, changing them would be a bad idea. Most desktop applications are designed for a non-linear ITF, and would look terrible, or at least very different, linearized. Therefore, the rendering pipeline, not the display hardware, must include the functionality to produce the desired appearance to the viewer.

Classic graphics texts propose displaying the linear intensities produced by the rendering system directly. However, research in the image reproduction domain by Bartleson & Breneman suggest that this will only give the best appearance if the viewing environment matches the original scene. If the viewing environment is darker, the reproduction will need more contrast to look right. They propose specific increases in contrast, expressed as gamma values, to accommodate dim and dark surrounding environments. These are used in the image reproduction industry to improve the appearance of television (dim surround) and projected film (dark surround). Work in web-based graphic standards such as sRGB recommend a non-linear viewing gamma on computer displays, and color appearance models such as CIECAM 97 use the Bartleson & Breneman results to model the effect of the surrounding illumination on color appearance.

Work by Jack Tumblin and Holly Rushmeier apply more complex visual models to determine the ITF for rendered graphics. They describe their results in terms of tone reproduction, and base their results on the Stevens models of brightness. In their results, and in other work by Tumblin, they can more accurately reproduce high-contrast imagery, and create renderings that reflect the absolute brightness of the image. That is, a room lit by a dim light still looks dim compared to one lit by a bright light. Simpler models would normalize both images to the same appearance. Clearly, including perceptual models as part of the rendering pipeline is an important part of making rendered images look more realistic.

Color Selection

One aspect of digital color familiar to almost all users of computer systems is the color selection tool. Color is applied to documents, desktops, drawings and web pages. A typical desktop system might have over a dozen different tools for selecting RGB colors embedded in its applications. I know of no textbooks about designing color selection tools. There are, however, a few ideas that consistently appear. First is the transformation of the RGB color cube into some approximation of a perceptual colorspace. The two most commonly used are the HSV and HLS systems described in the Foley and van Dam graphics texts. Second is the mapping of a 3D space to a 2D, interactive tool. The simplest mapping is sliders for each dimension such as R, G and B. This can be applied to any set of primaries, so graphic arts systems also include C, M, Y and K sliders. Many tools map two dimensions to a plane, then include a linear slider for the third. Many different mappings are possible, and the slides show several examples. Finally, color tools often include palettes of standard colors plus user-defined colors. The standard colors often represent the colors available on a display limited to 16 or 256 colors. There is also usually a pair of patches that show the color being changed and its new value.

One of the original applications of perceptually-based color systems like the Munsell book of color, or the Ostwald system, was for color selection. There have been some efforts to use perceptual spaces as the basis for digital color tools, as well. One of the first that was commercially available was “TekHVC” by Tektronics. The TekHVC space is similar to CIELUV, and the “TekColor Picker” could show slices of the gamuts of two different devices, making it possible to select only in-gamut colors. Giordanno Beretta’s

"MetaPalette" used the CIELAB space to select palettes of harmonious colors. His work influenced the "Canon Color Advisor," a clever tool buried in the print driver of Canon color printers.

Recent work at Brown University has resulted in the design of a suite of color selection tools based on design principles. These include user-defined palettes plus several innovative tools for creating and manipulating them. They allow the user to rough out the color composition before focusing on the form, and provide analysis tools for color frequency. This work will hopefully be published before the end of 2002.

Another form of color selection is to use color names. There is something fundamental about naming color. Ask someone to describe a color and they will begin first with a hue name such as blue or green. Berlin and Kay discovered that color names are common to human language, and that there are a standard set of 11 simple color names that appear in all mature languages. These are the perceptual primaries (red, green, blue, yellow, black, white and gray) plus orange, purple, pink and brown. Furthermore, the order the names are introduced into the language is consistent across languages. Boynton and Olson have further studied the psychophysics of color names with a set of color naming experiments. They found that there was a high degree of correlation between the names given colors across people and cultures. That is, the color patches described as "red" by one person will also be described as red by another. Of course, not all colors have simple names, and there is less agreement on these. But for the agreement for the basic colors was very good. A recent PhD thesis by Johan Lammens, which is available on-line, is an excellent reference for the perceptual basis for color names.

Color names are also important for industrial applications. In 1976, the National Bureau of Standards published *A Universal Color Language and Dictionary of Names*, by Kelly and Judd. This describes both an orderly language for specifying color, calibrated to the Munsell color space, and a mapping into this language of thousands of industrial color names. For example, the color "avocado" could be described as "dark olive green." Berk, Brownston, and Kaufman published an article in 1982 that introduced a simplified, more computationally tractable version of the NBS Color Language. We had an implementation of their system (plus our own improvements) running at Xerox PARC in the 80's. I found an advertisement for a commercial implementation of the NBS Color Language as part of the Color Science Library, from the CGSD Corporation in Palo Alto, CA (www.cgsd.com). It looks like a useful package, but I haven't tried it myself.

In the graphics arts domain, the Pantone Matching System® (www.pantone.com) is a common way of describing colors. Originally published as a book of print samples, the digital form of this system is incorporated in most of the major graphic arts applications such as Adobe Illustrator or Quark Express.

Finally, dictionaries of color names are starting to be used for specifying colors in HTML, X-Windows and probably other applications. The names I have seen are simply a binding of a text string to an RGB triple. I have some reservations about the names themselves (what color is "silver" anyway, on a computer monitor?) But, the idea of naming is a good one. Names tend to limit the set of colors specified, and this can be good, which is why specifying hundreds of names is bad. In the X-window implementation, the name dictionary provides a useful level of indirection in the color specification Say you design a tool that includes "brown" text on a "tan" background. You then discover that text becomes illegible with that combination of colors on a 256 color display. You can simply change the specification of the names once, and have the effect appear everywhere.

Color Design

I am not a formally trained graphic designer or illustrator, though I'm a pretty experienced amateur, as are many of us in the graphics field. For the discussion on visual design, I rely on a book by Wucius Wong called *Principles of Color Design.* Personally, I find Wong's systematic approach to design very clear and compelling. On the other hand, if you go to the graphic arts section of your local bookstore and look for a book on color design, most of what you will find is books of color schemes represented as color palettes. Whelan's, *Color Harmony 2* came with a palette selection tool (the PalettePicker) on CDROM, so I chose it as my example of such an approach. Edward Tufte is famous for his books on information visualization, and *Envisioning Information* contains a chapter on the use of color. I will conclude with a brief summary of his view on how and why to use color for visualization.

Color Harmony

The color circle used by designers is quite similar to the circle of hues described in a perceptually based color system such as Munsell. The circle is formed starting with three primaries, red, yellow and blue, plus three secondaries created by mixing the primaries pairwise: orange, green and purple. The full circle contains all the pairwise mixtures of these six colors as well. I am often asked why the primaries are not cyan, magenta and yellow. The answer may be more a one of naming than of function. The printing ink magenta is a slightly purplish red, and the cyan is more or a true blue than its overprint with magenta, which is a purplish, almost navy blue. To one not familiar with printing terminology, they could naturally be called blue, red and yellow.

The other reason my be simply one of history and application. The cyan, magenta and yellow primaries used in an image reproduction system are carefully chosen to create the largest possible gamut. They are subject to strong constraints as to function and cost. The designer's colorspace was traditionally formed by mixing paints. There is more flexibility in the exact choice of colors, which can be adjusted for an individual design. Whether layered or blended, the mixing principles are similar for both sets of subtractive primaries.

The other dimensions of a designer's colorspace are lightness and saturation. Again, these are similar to the perceptual spaces discussed earlier. In fact, Wong uses the Munsell space and terminology (hue, value and chroma) to describe color design. Knowing that Munsell was an artist as well as a color scientist, makes this less surprising.

The term *tint* applies to a color that has been lightened and desaturated by mixing it with white. A *tone* is a color grayed and darkened by mixing it with black. The *complement* of a color lies opposite it in the color wheel. A color mixed with its complement will become neutral, or gray.

The concept of *color harmony* is defined by Wong as "successful color combinations, whether these please the eye by using analogous colors, or excite the eye with contrasts."

Analogous colors are similar, contrasting colors are opposites. He then goes on to demonstrate these principles applied to the hue, value and chroma of colors, using many beautiful examples.

Color palettes are sets of 2-3 colors, plus their tints and tones, that are used in a design. They are often described in terms of qualities ranging from semi-scientific (contrast, neutral, primary) to emotional (warm, elegant, exciting). An important issue when designing a palette is the message you are trying to convey with your choice of color. Now, using palettes in design is an excellent idea. It provides an overall visual coherence that is very important. And, people's emotional response to color is very real. So whether or not you are comfortable with labels like "elegant," having access to predesigned palettes, or better yet, the ability to custom design effective color palettes, should help improve the appearance of our graphics. It is interesting, however, that all the color tools described in the previous section really only support selecting one color at a time. Even those that provide "palettes" do not provide any way to name, manage or reuse those palettes. Professional graphic design systems such as Adobe Illustrator seem to do a somewhat better job in this area, but I suspect there is far more that could be done to support palettes and other design principles in our digital tools.

Effective use of color

The goal in selecting colors is, ultimately, to produce an image that is both attractive and that conveys its message effectively. It is difficult to describe the "effective" use of color unconstrained by any application. However, there are a few common principles worth mentioning here.

Color is very eye-catching, creating emphasis, clustering, and evoking relationships to the known world. According to Tufte, the primary rule for the use of color "do no harm." Color should be clarifying, not confusing. It should be tasteful, and it should be robust across media, viewing conditions and viewers.

Color can be used to identify or label objects in an illustration. A classic example is shown in the slides. Take a page of numbers and count all the 7's. Studies have shown that the amount of time to count them is proportional to the number of digits on the page. That is, to count the 7's you must look at each digit in turn to determine whether or not it is a 7. Now, color the 7's some distinctive color such as red. Now the time it takes to count the 7's is proportional to the number of 7's on the page. The red color labels the 7's so all you have to do is count them up. Color works extremely well as this type of label as long as the identifying color is distinct within its context. Colored roadways or icons on maps, or colored text to identify hyperlinks are examples of color labels.

Color scales can be used to indicate quantity. There are no real intuitive hue orderings except possibly one for temperature (blue to red), but lightness or saturation scales are often effective for defining relative values. Colors distributed on a plane also create visual clusters of similar values. Examples are thermometer scales or density plots. Cynthia Brewer has a wonderful website illustrating her "Color Brewer" for creating color scales for visualization, as well as papers describing her theories. (http://www.personal.psu.edu/faculty/c/a/cab38/ColorSch/Schemes.html)

The world we see around us is colored, and colors in illustration often reflect their use in the physical domain. For example, stop signs and lights are red, so stop “buttons” are also. Geographic maps tint water blue and earth brown. The icons shown on the slide mimic the color of the fruits they represent, although the colors themselves do not match the natural object well. This is often true of color in illustrations.

Finally, we use color because it is beautiful. Most people spend some amount of effort to make the colors in their environment match their aesthetic taste. Look at the way people decorate their workspaces, or color their computer desktops. Color impacts emotions, which can impact performance. While the value of aesthetics is often questioned in “practical” disciplines like science and engineering, it impacts even these fields. An attractive illustration will be more appealing to look at and will therefore be more effective.

Adding color to illustrations and information visualization system in a way that is robust across systems, media and viewers can be very hard. Without color management, a carefully designed presentation on one medium can become a muddled mess on another. For example, I had a colleague who used bright green text on a dark green background for slides. On the monitor, the text was quite legible and the overall appearance was pleasing. Transferred to slides using a film recorder, however, the result was overall very dark and lacked the contrast needed to comfortably read the text. Even an excellent color management system would not have completely saved him, as his text color was far outside of the film gamut.

A primary issue for effectiveness is legibility, especially for text. They key to legibility is brightness contrast, which is why black text on white paper (or white text on dark slides) is so popular. In fact, an edge defined by only hue difference cannot be seen. Such an example is difficult to construct, but I can assure you, it is very disconcerting. You can see that two adjacent regions are differently colored, but it is impossible to find the edge between them. Another issue for text legibility is the distribution and focus of the short wavelength or blue cones as compared to the red and green cones. The blue cones are much more sparse, and focus on a slightly different plane than the other two. This is why some colors of bright blue text appear blurry—they are stimulating primarily the blue cones, which do not provide the same acuity as the others. The slides show a diagram from Mark Fairchild that shows the relative distribution of the three cone types. Note that there are no blue cones in the very center of the retina.

Another type of cross media problem has to do with grayscale and color-mapped displays. Grayscale systems, of course, must map all colors to gray values. Care must be taken that these grays will function as replacements for the colors. Many desktop systems display a maximum of 256 colors, using dithering to approximate the rest. Dithered text and other fine features are often illegible, so it is important to chose colors that will remain solid. This is also true for printers, especially for low-resolution digital desktop systems. Text and fine features are much more crisp when they contain at least one solid separation, even in offset printing. A rather nice feature of the Netscape web browsers is that they automatically map text to solid colors on dithered displays. I wish more applications did this.

All viewers are not the same either. As noted in the vision section, a significant part of the population has atypical color vision. As the red-green anomalies are the most common, one rule of thumb is to avoid encoding purely with a red-green distinction (an issue that apparently never occurred to the designers of stoplights). However, to ensure that important information can be distinguished by everyone, color should be combined with shape or other ways of encoding the information (like a stop sign). Another viewer issue is poor acuity, or lack of focus. Not all acuity problems can be fully corrected with lenses. Also, the eye gradually looses its ability to change focus as people age, which is why most people need some form of “reading glasses” by the time they are fifty. Avoiding small, low-contrast text and other features is important to accommodate this problem. Pill-bottle labels with small colored letters on a colored background should be banned!

In summary, when in doubt, use color conservatively. Duplicate the information carried by the color whenever possible. Know your target media and audience well, and be sure to test your designs thoroughly.

Annotated Bibliography

I have organized this bibliography by section, though there will naturally be some references that span sections.

Introduction

A good general book on color is Williamson and Cummins *Light and Color In Nature and Art.* It was designed for an undergraduate course and includes information on physics and optics as relates to color, as well as many of the topics covered in this tutorial.

1. S. J. Williamson and H. Z. Cummins. *Light and Color In Nature and Art.* John Wiley & Sons (1983).

Color Vision

There are many excellent color vision textbooks. I chose Wandell's because it is recent and has a computational flavor that should appeal to the technical members of the graphics audience. I also mentioned the text by Leo Hurvich, which is out-of-print but should be available in academic libraries. There are a couple of other vision texts worth mentioning, though I am personally only familiar with one of them. Cornsweet's 1970 book is a classic, though now 30 years old. David Marr's book provides a less conventional yet intriguing model for vision.

Wyszechki and Stiles massive *Color Science* is the standard reference for colorimetery and color psychophysics, though an introduction to these topics is included in almost any book on color. A substantial amount of this book is tables of numbers, which nowadays would be presented on CDROM. Interestingly enough, this book has no color pictures. Another classic book on color science and its application is Judd and Wyszechki's *Color in Business, Science, and Industry*.

Mark Fairchild's book is the best reference on color appearance models. Josef Albers book, *Interaction of Color*, is famous for its color illustrations. It comes in two forms, an expensive hardback ($135) with a full set of color plates, or a somewhat abridged paperback version for approximately $10. I have the paperback, and have found it quite useful.

2. Brian Wandell. *Foundations of Vision*. Sinauer Associates Inc. (1995).
3. Leo M. Hurvich. *Color Vision*. Sinauer Associates Inc. (1980).
4. Tom Cornsweet. *Visual Perception.* Academic Press (1970).
5. David Marr. *Vision.* W. H. Freemand, San Francisco, CA. (1982).
6. Judd, D. B. and G. Wyszecki. *Color in Business, Science, and Industry*. John Wiley & Sons (1975).
7. G. Wyszechki and W.S. Stiles. *Color Science, Second Edition.* John Wiley & Sons (1982).
8. Mark D. Fairchild. *Color Appearance Models*. Addision-Wesley (1998).
9. Josef Albers. *Interaction of Color.* Yale University Press (1963).

Color Reproduction and Management

Robert Hunt's *The Reproduction of Colour* is an invaluable reference for color reproduction technology. Now in its 5th edition, it is a must-have for anyone interested in this aspect of color. Gary Field has two newer books on the topic, *Color and Its Reproduction, 2nd Edition* and *Principles of Color Reproduction*, an update of John Yule's classic book on printing, which has long been out of print. Giorgianni and Madden's *Digital Color Management* is a well-written book with a strong engineering orientation. It is beautifully produced, printed in full color, with some of the nicest illustrations around. In spite of the full-color production, it is reasonably priced at around $70. To understand color management from a user's perspective, I recommend either *Understanding Digital Color* (another book with outstanding illustrations) or *Adobe Photoshop 6.0 for Photographers.* Even if you are not a Photoshop user, it's description of RGB color spaces for image manipulation is worth looking at. Charles Poynton, who regularly teaches about digital color at SIGGRAPH, has written two references on digital video. His courses often include a section on color reproduction and color management. There is active research in the area of digital color reproduction systems and color management. Much of this work is published at conferences and in publications supported by the *Society for Imaging Science and Technology*, or IS&T (www.imaging.org). Their "Recent Progress" Series are bound volumes of select papers on topics such as color management and color science. The Graphic Arts Technical Foundation (www.gatf.org) is also a good source of practical information on color and printing. Its publications can be ordered online through the Graphic Arts Information Network (www.gain.org). The Society of Information Display (www.sid.org) is a good source of references on display technologies. I've found *Display Systems* and *Projection Displays*, from the Wiley SID Series in Display Technology particularly useful.

10. R.W.G. Hunt. *The Reproduction of Colour, Fifth Edition.* Fountain Press, England (1996).
11. G. G. Field, *Color and Its Reproduction, 2nd Edition,* GATF Press, 1999.
12. J. A. C. Yule with G. G. Field, *Principles of Color Reproduction: Applied to photomechanical reproduction, color photography, and the ink, paper, and other related industries*, GATF Press, 2001.
13. E.J. Giorgianni and T. E. Madden. *Digital Color Management.* Addison-Wesley (1998).
14. P. Green, *Understanding Digital Color.* Graphic Arts Technical Foundation, 2nd edition (June 1999).
15. M. Evening, *Adobe Photoshop 6.0 for Photographers*, Focal Press, (January 2001).
16. C. A. Poynton. *A Technical Introduction to Digital Video*. John Wiley & Sons (1996).
17. C. A. Poynton. *Digital Video and HDTV Algorithms and Interfaces.* Morgan Kaufmann Publishers, (July 2002).
18. L.W. MacDonald and A.C. Lowe, *Display Systems,* John Wiley and Sons (1997).
19. E.H. Stupp and M. S. Brennesholtz, *Projection Displays,* John Wiley and Sons (1999).

20. M. Stone. "Color and Brightness Appearance Issues in Tiled Displays," *Computer Graphics and Applications,* 21(6), September 2001, pp. 58-67.

Color Synthesis

The primary reference I used for the rendering section of this tutorial is Andrew Glassner's big, two-volume work, *Principles of Digital Image Synthesis.* I am not a rendering expert, so I relied heavily on Andrew's presentation. Andrew's references are encyclopedic, but I duplicate a few here for convenience. Roy Hall's book is out of print and not as recent, so I did not use it as much. Roy does have, however, a recent paper on the topic of spectral representations for rendering in the July 1999 issue of CG&A. Murphy and Doherty's beautiful book on the forms of color in nature can be ordered through the Exploratorium website (www.exploratorium.com). The McCamy reference is for the Macbeth ColorChecker used in the linear models example, which can be found at many photography stores or through GretagMacbeth (www.gretagmacbeth.com)

21. Andrew Glassner. *Principles of Digital Image Synthesis.* Morgan Kaufmann, San Francisco, CA (1995).
22. Gary Meyer. "Wavelength Selection for Synthetic Image Generation," *Computer Vision, Graphics, and Image Processing*, vol. 41, pp. 57-79. (1988).
23. Mark Peercy. "Linear color representations for full spectral rendering," Computer Graphics, 27, 3, 191-198, August 1993.
24. Carlos Borges, "Trichromatic Approximation for Computer Graphics Illumination Models," *Computer Graphics*, 25,4, July 1991, pp. 101-104 (1991).
25. Paul Strauss, "A realistic lighting model for computer animators." *IEEE Computer Graphics & Applications*, 10(11):56-64, November (1990).
26. Roy Hall. *Illumination and Color in Computer Generated Imagery.* Springer-Verlag, NY, NY. 1989.
27. Roy Hall. "A Comparison between XYZ, RGB, and Sampled Spectral Color Computation." *IEEE Computer Graphics and Applications,* July, 1999.
28. P. Murphy and P. Doherty, *The Color of Nature,* Exploratorium Books (1996).
29. C. S. McCamy, H. Marcus, and J. G. Davidson. "A color-rendition chart." *Journal of Applied Photographic Engineering,* 2(3):95-99, summer (1976).

Artistic, or non-photorealistic rendering has been an expanding topic in graphics since the mid-90's. Winkenbach's work on applying Salisbury's pen and ink textures to rendering is one of the first in this series applied directly to 3D rendering. Curtis extends the model to watercolor. The original Kubelka-Monk reference is difficult to find, but Glassner includes a good explanation in his book. Gooch shading, presented at SIGGRAPH 98, extends the Phong model to techniques for technical illustration. The Gooches have a written a book on the topic.

30. M. P. Salisbury, S. E. Anderson, R. Barzel and D. H. Salesin. "Interactive Pen-and-Ink Illustration," Proceedings of SIGGRAPH 94, (Orlando, FL July 24-29, 1994), in *Computer Graphics Proceedings*, ACM SIGGRAPH, pp. 101-118 (1994).

31. G. Winkenbach and D. H. Salesin. "Computer Generated Pen-and-Ink Illustration," Proceedings of SIGGRAPH 94, (Orlando, FL, July 24-29, 1994), in *Computer Graphics Proceedings,* ACM SIGGRAPH, pp. 91-100 (1994).
32. C. Curtis, S. E. Anderson, K. W. Fleischer, D. H. Salesin. "Computer-Generated Watercolor," Proceedings of SIGGRAPH 97, (Los Angeles, CA, August 3-8, 1997), in Computer Graphics Proceedings, ACM SIGGRAPH, pp. 421-430 (1997).
33. A. Gooch, B. Gooch, P. Shirley, E. Cohen. "A Non-Photorealistic Lighting Model for Automatic Technical Illustration," Proceedings of SIGGRAPH 98, (Orlando, FL, July 19-24, 1998), in *Computer Graphics Proceedings*, ACM SIGGRAPH, pp. 447-452 (1998).
34. A. Gooch, B. Gooch, *Non-photorealistic rendering,* A.K. Peters, Natick, MA, (2001).

I first explored color issues with the exchange and rendering of 3D graphic for the web and desktop environments working with the VRML standard (www.web3d.org). A paper at the 1997 Color Imaging Conference describes it as a "work in progress." Charles Poynton has written extensively about gamma, both formally and on his website (http://www.inforamp.net/~poynton/). The sRGB colorspace and recommended practices have been adopted by the International Electrotechnical Commission (IEC) as IEC 61966-2-1 (http://www.iec.ch/home-e.htm).

35. Maureen Stone. "Color in Web-based 3D Graphics," Proceedings of the IS&T/SID Color Imaging Conference, Scottsdale, AZ, November (1997).
36. Charles Poynton. The Rehabilitation of Gamma," Proceedings of SPIE 3299 (San Jose, Calif., Jan. 26-30, 1998), 232-249 (Bellingham, Wash.: SPIE, 1998).

Tumblin and Rushmeier's original paper is available both as an IEEE CG&A journal article and as a tech report from Georgia Tech. Tumblin has done further work in this area, some of which is reflected in the TOG article listed, his SIGGRAPH 2000 paper, and in his PhD thesis. Mark Fairchild, together with the Cornell computer graphics group, have been working to develop a framework for appearance and rendering, as was reported in a SIGGRAPH '98 paper.

37. J. Tumblin & H. Rushmeier, "Tone Reproduction for Realistic Images" *IEEE Computer Graphics & Applications,* Vol. 14, No. 6, pp.42-48. Nov. 1993. Extended versions as Ga. Tech GVU Center Technical Report GIT-GVU-91-13, and 92-32.
38. J. Tumblin, J. Hodgins, & B. Guenter. "Two Methods for Display of High Contrast Images" *ACM Transactions on Graphics,* **18,** 1, January 1999.
39. S. N. Pattanaik, J. Tumblin, H. Yee, D. P. Greenberg, "Time-Dependent Visual Adaptation for Realistic Image Display," Proceedings of SIGGRAPH 2000, (New Orleans, LA, July 23-28, 2000), in *Computer Graphics Proceedings*, ACM SIGGRAPH, pp. 47-54 (2000).
40. S.N. Pattanaik, J. A. Ferwerda, M.D. Fairchild, D.P. Greenberg. "A Multiscale Model for Adaptation and Spatial Vision for Realistic Image Display." Proceed-

ings of SIGGRAPH 98, (Orlando, FL, July 19-24, 1998), in *Computer Graphics Proceedings*, ACM SIGGRAPH, pp. 287-298 (1998).

I know of no scholarly references on designing color selection tools, so I relied on my own observations, which I hope are not too idiosyncratic. The HSV colorspace was invented by Alvy Ray Smith and presented in a 1978 SIGGRAPH paper. The following year the HLS space was recommended by the Computer Graphics Standards Planning Committee. These spaces are also described in the Foley and van Dam textbooks. Similar transformations of the RGB color cube have probably been created by many practitioners. The TekHVC space and the TekColor Picker were developed by Tektronics in the late 1980's, as was the work at Xerox on MetaPalette. The MetaPalette evolved into the Canon Color Advisor, which was described in at a recent IS&T Color Imaging Conference. There is recent work at Brown University on color selection tools based on both perceptual and artistic principles. This work will hopefully be published soon.

41. Alvy Ray Smith "Color gamut transform pairs," *Computer Graphics*, 12(3):12-25. (1978).
42. "Status Report of the Graphics Standards Planning Committee," *Computer Graphics* 13(3), August (1979).
43. J. Foley, A. van Dam, S. Feiner and J. Hughes. *Computer Graphics Principles and Practices*. Addison-Wesley (1990).
44. J. M. Taylor, G. M. Murch and P. A. MacManus. "Tektronix HVC: A Uniform Perceptual Color System for Display Users," *SID 88 Digest*, pp.77-80, May (1988).
45. G. Beretta. "Color Palette Selection Tools," SPSE's 43rd Annual Conference, The Society for Imaging Science and Technology, 20-25 May 1990, Rochester (New York, USA) pp. 94-96 (1990).
46. L. Lavendel and T. Kohler. "The Story of a Color Advisor," Sixth IS&T/SID Color Imaging Conference (Scottsdale, AZ, November 1998), pp: 228-232. (1998).
47. B. J. Meier, A. M. Spalter, D. B. Karelitz, R. M. Simpson. "Interactive Color Palette Tools" Brown University, Department of Computer Science. Planned publication route: Journal of Graphical Tools.

Color naming has been well studied. The most complete reference in the psychophysical domain is Lammen's thesis. It contains many other references in its bibliography. I include the reference to Boynton and Olson's work here also. Berlin and Kay's anthropological work on color naming was originally published in 1969, and was reprinted in 1990. The National Bureau of Standards has become NIST, the National Institute of Standards and Technology (www.nist.gov). The Pantone color matching products can be ordered from Pantone (www.pantone.com).

48. Johan M. Lammens. "A Computational Model of Color Perception and Color Naming," Ph.D. dissertation, State University of New York at Buffalo, Computer Science department. (1994).
49. Robert Boynton and Conrad Olson. "Locating basic colors in the OSA space," *Color Research and Application*, 12(2):94-105. 1987.

50. Brent Berlin and Paul Kay. *Basic Color Terms: Their Universality and Evolution,* University of California Press, Berkeley CA, 1991 edition., originally published in 1969.
51. K.L. Kelly and D. B. Judd. *Color, Universal Language and Dictionary of Names*, National Bureau of Standards (U.S.) Special Publication 440, Washington, (1976).
52. T. Berk, L. Brownston, and A. Kaufman. "A new color-naming system for graphics languages." *IEEE Computer Graphics and Applications*. 2, 3 (1982).

Wong's books on design are systematic and elegant in a way that should appeal to many in the computer graphics field. His *Principles of Color Design* is based on the Munsell (www.munsell.com) color ordering system. Whelen's *Color Harmony* contains a full set of palettes plus a Windows application called the "Palette Picker" that is also available from www.lightdream.com. The Palette Picker lets you select palettes in three ways: by "scientific" name, by "aesthetic" name, or to harmonize with a selected color. The tool is not bad—it comes up with plausible palettes. It can only display their RGB or CMYK encoding, however, which then have to be copied by hand into another application. Tufte is famous for his books on information visualization. The only one of his three books that contains a chapter about color is *Envisioning Information*. Colin Ware has written a book about visualization and perception. He includes several chapters on color and related topics. Cynthia Brewer, who is a professor in the department of Geography at Penn State, has worked extensively with techniques for creating color scales. Her papers and website are a good resource for those interested in color scales for map data.

53. Wucius Wong. *Principles of Color Design, second edition.* John Wiley & Sons, NY, 1997.
54. Bride M. Whelan, *Color Harmony 2.* Rockport Publishers, Rockport, Mass., 1997.
55. Edward R. Tufte, *Envisioning Information.* Graphics Press, 1990.
56. Colin Ware, *Information Visualization, Perception for Design.* Morgan Kaufmann, San Francisco, CA, 2000.
57. Brewer, Cynthia A., 1994, "Color Use Guidelines for Mapping and Visualization, Chapter 7 (pp. 123-147)" in *Visualization in Modern Cartography*, edited by A.M. MacEachren and D.R.F. Taylor, Elsevier Science, Tarrytown, NY.
58. Cynthia Brewer: http://www.personal.psu.edu/faculty/c/a/cab38/

A Field Guide to Digital Color: Color at Rhythm & Hues

Pauline Ts'o
Rhythm & Hues

Introduction

Rhythm & Hues is a high-end computer animation and visual effects studio. We create effects and animation primarily for feature films, special venue films, and commercials. Our work ranges from all *cgi*, computer generated imagery, which may be quite stylized, to absolutely photorealistic live-action and cgi combination work.

We will review some of the basic aesthetic goals that are specific to the 3D work at Rhythm & Hues, as well as the main color management processes for video projects and for film projects.

Aesthetic Goals

Color is only one part of the cinematographer's toolkit. Let's first consider the general goals of cinematography in film-making. The overall goal of cinematography is effective narrative by 1) directing the eye to specific and important storytelling visual elements and 2) creating an overall visual mood. Traditional cinematography has two fundamental components with which to achieve these goals - 1) the capture of light by a camera lens onto film or video and 2) the motion of the camera lens through 3D space.

Let's briefly consider how one directs the eye. By far, the most effective way to do so is through contrast. Our brains are wired to be particularly sensitive to contrasts of all types - motion, texture, sharpness, size, etc., - as well as color contrasts in hue, saturation and value. (See Figure 1, at the end of the notes.) Our brains are attracted to differences.

In 3D film-making, as opposed to 2D graphic design, the role of contrasts of hue and saturation is often diminished. This is at least in part because the camera is moving through 3-dimensional space, and thus shadows (contrasts of value); depth-of-field (contrasts of sharpness); and contrasts of motion are added to the storytelling equation. Shadows and depth-of-field, in particular, fall uniquely out of dealing with 3-dimensional space.

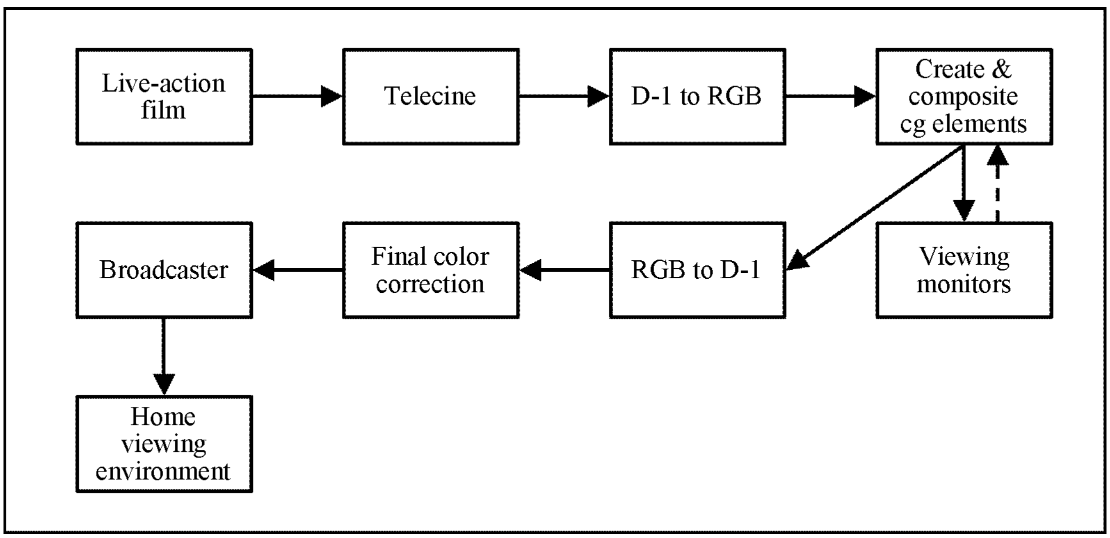
In fact, as an art and lighting director for 3D animation, I have found that contrast in value (or brightness, i.e., the achromatic component of color) is the most important variable to consider. It is not unusual for me to view an image in gray scale to check the composition and balance of the lights and darks. This is not to say the other types of contrast are unimportant. Indeed, they should all be approached consistently to avoid visual confusion.

Hue and saturation play a much more prominent role when trying to create an overall visual mood. Different palettes evoke different emotional responses that have evolved both from shared experiences in the natural world, and perhaps not as shared experiences within different cultural worlds. These emotional responses are very subjective. (See Figure 2.)

Visual effects for live-action often have an additional aesthetic goal that graphic designers and traditional cinematographers are inherently less worried about - recreating believable photo-realism from nothing. How does color factor into this goal? Oftentimes our computer generated, or *cg*, element has to match an existing live-action element as well as be integrated into a live-action plate. Here, color matching of materials and lights between the live-action elements and the cg elements becomes critical and much less subjective. (See Figure 3.)

Video Color Pipeline

For television commercials, our final delivery medium is D-1, a digital, non-linearly encoded, YC_BC_R component videotape format (one luminance and two chrominance components). Because our final output is digital and the final viewing medium, the television, is pretty much the same as our previewing medium, computer monitors, our color pipeline for video is relatively straightforward. The following is a summary of the major steps where color may possibly be altered.



The dashed lines indicate when the results of a step may show that an earlier step needs to be revisited, not that the output of a step serves as actual data for an earlier step. We'll now explore each of these steps.

Live-action Footage

If the commercial has live-action that is to be combined with cg elements, the live-action is first shot onto film with all the usual traditional cinematography considerations, then transferred to D-1 videotape via a standard *telecine* session. A telecine is a device that essentially uses a film projector and a scanner or camera to convert film to video and is also capable of performing color correction. These sessions occur off-site and so the technical details of their color management pipeline are not part of the expertise at Rhythm & Hues.

Next, we use a Discreet Logic Flame or Inferno system (www.discrete.com/products) to read the data off of tape and onto disk. From there, we run a proprietary program which converts the video encoding to linear intensity, RGB space, and stores the converted data to a proprietary file format called *rll*.

File Format

We use the same rll format for both film and video. This file format is a 16-bit, floating point format. That is, for each of the red, green, blue, and matte channels, 16 bits of data per pixel is stored. Many common file formats use only 8-bits of data per channel. Eight bits of data per channel is often good enough for video resolution, but there are times when it is not and visible banding will occur, especially in the dark blues. Banding problems can also occur within an otherwise 16-bit environment when using a texture map that has been created in an 8-bit environment such as Photoshop. When banding occurs, dithering is one way of diminishing its effect. Another much less desirable way to address banding is to actually change the colors.

While running a Rhythm & Hues image program, images are manipulated and stored in memory using 32 bits of IEEE format floating point data per pixel per channel. However, storing an image onto disk with this bit depth resolution is costly. Simple run-length encoding is not very useful, especially with live-action footage, given the amount of high-frequency detail commonly present. Therefore, we use another type of encoding where the entire possible set of 32-bit floating point numbers in memory is mapped into a smaller subset of 16-bit numbers and these are stored to disk instead. Each of these 16-bit numbers is called a *code*.

There are many ways to map an evenly spaced set of floating point numbers to the codes. One common way is *linearly*, where the codes representing these floating point numbers are also evenly spaced. Another way is *logarithmically*, where codes representing darker values are closer to each other than the codes representing brighter values. Log encoding prevents visible banding in dark areas where subtle changes in color may occur. This is the mapping that the rll and 10-bit Cineon format uses. Lastly, a non-linear mapping based on an exponent of 1/2.2 is common for many 8-bit file formats. This is controversially sometimes called "gamma 2.2 corrected". The following equations and formulas might help explain.

If

I_{min}	is the dimmest intensity
I_{max}	is the brightest intensity
N	is 1 less than the number of codes
I	is the intensity
C	is the code number

then

linear

$$C = N * (I - I_{min}) / (I_{max} - I_{min})$$

gamma 2.2 corrected

$$C = N * ((I - I_{min}) / (I_{max} - I_{min}))^{0.454545}$$

logarithmic

$$C = N * (\log(I) - \log(I_{min})) / (\log(I_{max}) - \log(I_{min}))$$

Here is are tables showing the codes resulting from these formula:

	I_{min}	I_{max}
Rll code	5.877×10^{-39}	6.779×10^{38}
8-bit linear code	0.0	255.0
8-bit gamma corrected code	0.0	255.0
Cineon code	.005208	13.386488

Floating point value (I)	0.003906	0.007813	0.015625	0.03125	0.0625	0.125	0.25	0.5	1.0
Rll code (0-32767)	15232	15360	15488	15616	15744	15872	16000	16128	16256
8-bit linear code (0-255)	1	2	4	8	16	32	64	128	255
8-bit gamma corrected code (0-255)	21	28	39	53	72	99	136	186	255
Cineon code (0-1023)	0	53	143	234	324	414	504	595	685

The number of possible codes for each format are listed in parentheses in the second table. For the rll format, the negative range of possible codes is reserved for other uses. Notice that I increases by a factor of two from column to column, thus representing a doubling of the amount of light, or one *stop*. Note also that given these stops, the rll and Cineon codes are evenly spaced. For a visual feel of the difference between linear and log encoding, see Figure 4. For more information on the Cineon format, see: www.cinesite.com/CineonTech/GrayScale/GSv2.html.

For our rll format, the floating point value 1.0 was mapped to the code number 16256 for computational efficiency reasons relating to the particulars of the IEEE floating point format. The Cineon format I_{min} and I_{max} were chosen with the idea of adequately capturing the dynamic range of typical motion picture negative film. So, for example, the Cineon I_{max} provides about 3.5 stops of latitude over Kodak's standard white. Our rll format admittedly provides far more dynamic range representation than is practically useful, allowing for 128 steps between one stop and the next full stop, much more resolution than the human eye can distinguish.

To complete the rll encoding algorithm, it is the high 16 bits of the 32-bit floating point number which are mapped logarithmically to the codes. Once an image is mapped to this subset of 16-bit numbers, run-length encoding and other encoding techniques are also used to reduce file size.

Viewing Monitors

Cg elements are then created via the usual steps of modeling, animation, rendering, and compositing. During this part of the pipeline, rendered and composited images are being regularly viewed by digital artists on various computer monitors. These monitors, therefore, need to be calibrated so that everyone is seeing the same thing.

Using a Minolta color analyzer to read the monitor output, we use first the brightness and contrast knobs on the front of the monitor and then the built-in bias and gain controls to set the monitor white point to D65, corresponding to standard daylight at 6500° Kelvin color temperature, and the maximum luminance to be about 30 foot-lamberts. A dark point is also similarly calibrated to be 0.3 foot-lamberts. The color analyzer contains 3 photocells and 3 filters that approximate the response of the cones in the eye, i.e., an approximation of CIE tristimulus color space. Digital artists are then vigorously encouraged not to touch the brightness and contrast controls on their monitors after calibration.

Video images are pretty consistent with respect to color across calibrated monitors as the phosphors typically used by various manufacturers are similar. Monitors, however, do drift out of calibration, so periodic hardware re-calibration is important. Often, it will be a lighting director or cg supervisor who notices that a particular monitor needs calibration because the directors and supervisors visit many monitors and basically need to develop the visual equivalent of perfect pitch.

Hardware calibration eliminates one set of viewing variables. However, there are plenty of other variables. One such is the viewing environment. While the phosphors of a monitor are light-emitting, the glass covering the phosphors is often quite reflective. Thus, environment lighting in the work place can have a significant effect on perceived brightness and contrast of images displayed on monitors. The digital artists who work on rendering and compositing at Rhythm & Hues therefore usually work in dimly lit areas.

Controlling Gamma

Another variable is operating system gammas. We will define an *OS gamma* as the net effect of the monitor's gamma function and any "gamma correction" assumed by the operating system or applied by a program such as Silicon Graphics' ("SGI") gamma program. Gamma corrections can be applied by an operating system or by a software application. Within the Rhythm & Hues family of software running on SGI machines, an OS gamma of 2.2 is assumed. That is, if a user were to simply type "gamma" on an SGI machine, he or she should expect that the value returned by that program will be 1.0 (the number being returned by the SGI gamma program is actually the inverse of the exponent used in the power function describing the monitor's gamma function). Another way of thinking of this is that all of Rhythm & Hues' rendering and compositing tools use a linear intensity space while performing computations. Different gamma and color corrections are then applied by Rhythm & Hues display programs depending on whether the image is meant for film or video output. Some third-party software tools, however, assume an OS gamma of 1.0. That is, if a user were to type "gamma" on an SGI machine, the number returned would be 2.2. This can lead to confusion and the digital artists must be mindful of the different assumptions.

When dealing with video, our display tools apply a software gamma correction of 2.2 to the raw image data. The data is also clipped for display to be within the range of 0.0 and 1.0.

Last Steps

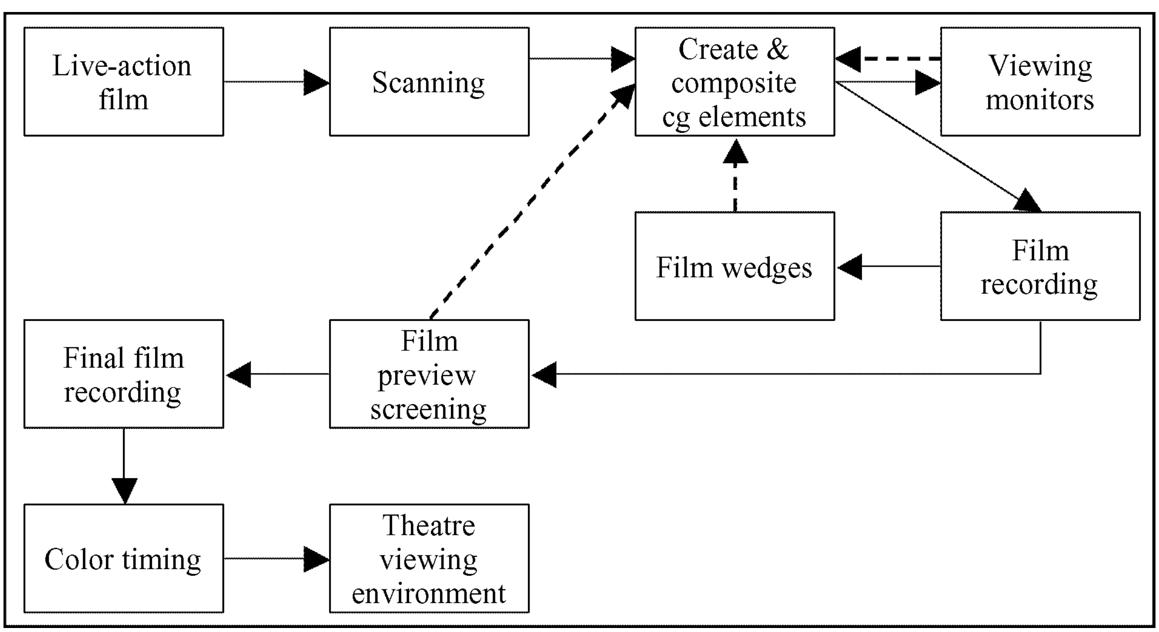
Once the images are finalled using the calibrated monitors at Rhythm & Hues, the images are to be recorded onto D-1 tape. RGB data is converted to YC_BC_R data and a non-linear mapping is applied. There is one more check to be made before the final record. While all possible D-1 component values can be processed by most video display devices, when the broadcaster converts these component signals into a single NTSC composite signal that is sent to the home TV, the peak composite waveform may exceed certain values which are then clipped, causing a possible color shift. Yellows are usually the problem area. At Rhythm & Hues, we have the ability to preview the composite signal on a waveform monitor.

From a Rhythm & Hues-produced D-1 tape, the images may be taken to a final off-site session where overall last-minute color rebalancing can be made according to the client's whim. A common system used would be a da Vinci color corrector (www.davsys.com/product_groups.htm). The output of this session is usually another D-1 tape.

From this final D-1 tape, the images are converted into the composite signal which is broadcasted to the viewing audiences at home. Of course, at this point, it is unknown how these televisions are set for color, brightness and contrast, and under what lighting conditions the images are viewed.

Film Color Pipeline

The film color pipeline is much more complicated than the video color pipeline because 1) previewing on film is slow, expensive and cumbersome, and 2) faster previewing on a video monitor is not an equivalent experience. The following is a summary of the major steps where color may possibly be altered.



Again, the dashed lines indicate when the results of a step may show that an earlier step needs to be revisited, not that the output of a step serves as actual data for an earlier step.

Scanning

If the film project involves combining cg elements with live-action elements, then the first step is to digitize the live-action negative using a custom CCD scanner and the same proprietary rll file format used for video.

Our scanning department hopefully receives a *match clip* from the client. A match clip is a short, 3 frame piece of the work print, i.e., positive film, which has been approved by the client for look. "Look" encompasses mainly color, brightness, and contrast. The match clip is then scanned in.

The portion of the negative which corresponds to the frames of the match clip is then also scanned in and manipulated to look like the match clip on a monitor by changing the amount of red, green, and blue exposure. This is done via human eyeballs. The rest of the negative is scanned using these RGB adjustments.

Then, a single frame from the match clip is shot to film by our laser film recorder as a 27-step *wedge*. By a "27-step wedge" we mean a series of 27 variations of the single frame is created by modifying the amounts of red, green, and blue. One of these 27 frames will be identical to the digitized single frame from the match clip, i.e., no RGB modification will have been performed on it. (See Figure 5.)

The next step is to have the film lab print the wedge using *LAD standards*. LAD stands for "Laboratory Aim Density". A film lab processes a lot of footage and as a cost-savings measure, they try to use the same chemical print bath for as long as possible. However, this means that prints can noticeably vary in appearance.

With LAD printing, we shoot, along with the regular footage, a frame of 16% grey (a Kodak standard) and another standard frame which has various color patches and an image of a woman's face. The film lab uses the grey patch to calibrate their printer using a light measuring device called a *color densitometer*. The densitometer measures the film's *density*, which is the negative log base 10 of its transmittance. If the film lab's grey patch is within +/- 0.08 units of brightness and +/- 0.06 units of red, green, and blue of our status A density target or *aim* (which is (1.10, 1.06, 1.03)), then the print is considered to be a "good LAD print".

Once our 27 frame wedge comes back from the film lab as a good LAD print, the effects supervisor will place the wedge on a light table next to the original match clip frames and pick which of the wedge frames looks most like the match clip frames. Based on this pick, the scanned sequence will have the same red, green, blue modifications applied and the sequence is then considered to be "color corrected".

So even before the digital artists are given access to the files, this color correction process essentially conforms our film scanning and recording devices to the expectation of the client. Remember that the live-action match clip could originate from any of various different types of film stock with additional color-timing or other chemical film processes tossed in. It would thus be impossible for our scanning and recording environment to anticipate and pre-model every possible scenario that could generate a particular match clip.

Once the live-action footage has been scanned and converted to rll files, or if there are no live-action elements, computer generated elements are then again created via the usual steps of modeling, animation, rendering, and compositing.

Rendering

Within the process of rendering, there are some color issues to be aware of. The first is to note that at Rhythm & Hues, we represent color in the rendering program as standard RGB triples, rather than spectral representations, primarily for historical reasons.

Another is assigning a *white point*. Any white point, whether we are speaking of monitors, film, or file formats, is an arbitrarily chosen definition for what is white.

Film has a reduced response (meaning that small differences are lost) near its maximum recordable brightness, and therefore, it is not advisable to depend on accurately creating values in that range. The solution is to arbitrarily pick another point that is a fairly bright white and define that point as white. At Rhythm & Hues, we now use the same white point that Kodak's Cineon file format assumes. So, our color range is numerically represented in our rll format as floating point numbers "between" 0.0 and 1.0, with our white point of (r,g,b) = (1,1,1) being mapped to a film negative density of 1.37, relative to the base density. This is a standard white point defined by Kodak (www.cinesite.com/CineonTech/GrayScale/GSv2.html). This means that values greater than 1.0 actually have meaning and are therefore not clipped. These greater than 1.0 values allow specular highlights and other bright light effects to be represented. (See Figure 6.)

Aside from the increased color resolution that comes with the 16 bit file format, the rll file format also allows negative values. Negative color values are ultimately clipped (by the film recorder), but the ability to have negative values means that intermediate computational steps in rendering and compositing retain as much information as possible until the very end. Also, negative values allow us to optionally store other types of data, such as depth information, within the same file.

Image Display

During this part of the pipeline, rendered and composited images are being previewed on the same calibrated monitors as are used when viewing video images. However, our image display software is notified that the intended output will ultimately be film and the software performs color correction accordingly. We use two different types of color correction, one we term "film", the other we term "quick film". These are not industry standard terms.

We have previously discussed the video monitor's non-linear response curve, which was described as the monitor gamma function. Film has its own non-linear response to light, which we will call the *film gamma*.

"Film" color correction involves both a film gamma adjustment and a color adjustment through an empirically derived look-up table. The look-up table was created by shooting color patches to film, directly measuring the film with the Minolta color analyzer and a white light (a tungsten light source corrected with standard filters to daylight color temperature), then using the known monitor characteristics to mathematically model the color space for the look-up table. The film gamma adjustment can be thought of as being applied in addition to the adjustment for the monitor gamma, although it is done in a single computational step. “Quick film” only applies a film gamma adjustment, and not a color adjustment. The film gamma adjustment used in quick film was derived from gray patches which were a subset of the color patches used to create the look-up table. It is therefore quicker to compute and used when time is of the essence. (See Figure 7.)

There are times, however, when digital artists will purposefully look at the composited live-action and cg elements using an OS gamma that is drastically higher than normal.

Why do we do this? Because no matter how good our color correction algorithms might be, monitors do not have the same dynamic range that film does. In particular, dark values that look the same on a video monitor might record to film at significantly different levels. Nothing gives away the fact more that a digital element has been added to a live-action plate than when the value range of the digital shadows doesn’t match the value range of the live-action shadows. Therefore, by boosting the OS gamma, we exaggerate differences within the low end of the dynamic range and they become more apparent when viewed on a video monitor.

Our display tool also has an option in which it indicates that the computed values are outside the color gamut of the monitor and have been clipped for display. In the cases where the color mapping is unavoidably displayed inaccurately, it is the responsibility of the visual effects supervisor or art/lighting director to create the correct mapping in his or her mind and transmit this understanding to the digital artists.

Texture Maps & Matte Paintings

Texture maps and matte paintings deserve a note here (see Figure 8.). Most commercial paint programs like Photoshop, Amazon and Studio Paint store images in 8 bit formats, such as Targa or tiff, which is often insufficient color resolution for film use. These file formats also are designed for a monitor and gamma correction which isn’t explicitly encoded in the image format. That is, a user may specify a gamma correction value, create an image using that value so that it looks “correct” within that user’s environment, but the file format of the image does not store what that gamma correction value was. Therefore, the interpretation of what the user’s OS gamma was is a guess at best. These programs also do not treat images meant for film any differently than images meant for video. So, if one were to shoot an image painted within one of these packages directly to film, the contrast and colors of the filmed image would look significantly different than the image as displayed by the paint package.

We address these issues in multiple ways. For 8 bit texture maps and matte paintings, we run a conversion program that tries to remap the data to linear space by assuming a non-linear encoding of 2.2 has already been applied. At this point, if the user were to view the texture map or matte painting with a Rhythm & Hues display program in video mode, the image would appear as expected, i.e., as it appeared on the user’s original monitor. However, in film mode, the image would not appear as expected because film’s color space is different from video's. Therefore, the same program essentially applies a film gamma adjustment (separate from the monitor gamma correction as discussed above) and color adjustment in reverse so that when the images are put through our film pipeline and the film gamma adjustment and color adjustments are applied then, the images will come out as expected.

Also, we usually store texture maps as 8 bit linear data to save on memory usage, but we can override this and specify 16 bit log or 32 bit float storage if necessary. Lastly, we have created a version of Gimp, an open source paint program very similar to Photoshop (www.gimp.org), that is 16 bit. Using Gimp is particularly important when applying paint touch-ups directly to a final rendered frame.

Recording to Film

Once we think that the images are pretty close to what we want, we shoot them out to film with an Arrilaser film recorder (www.arri.com/entry/products.htm) and project the film with a Kinoton FP30EC xenon arc film projector (www.kinoton.com/iprodukte.htm) onto a flat, matte white, gain 1 movie screen in a dark theatre. Again, no matter what we do, video monitors can never display film completely accurately. Therefore, actual film output is the only way to truly know what we have. However, film is expensive to shoot and turnaround for developing is slow. Therefore, we try to come as close as possible in video before shooting film tests. If we are matching a cg element to a corresponding live-action element, then we can usually come fairly close on the video monitor.

However, sometimes the only reference we have of the live-action element has been shot under different lighting conditions than the live-action footage that we have to work with. Then the choice of what colors create believable consistency becomes much more subjective.

Likewise, there are times, especially for special venue films, where the images are entirely computer generated. In these cases, we often shoot yet more wedges of color, brightness and contrast. Again, these wedges are printed to LAD standards and reviewed on a light table. Here, the "right" wedge pick is totally subjective and it is usually up to the art director or lighting director to make the appropriate decisions.

We eliminate one possible variable by always shooting our elements to the same film stock - Kodak 5242. Traditional cinematographers, of course, use many different types of film stock, and the way we match to these is by adding our own digital grain to our elements. Matching the color bias and contrast of different types of film stock falls out naturally from the normal process of rendering and compositing.

The calibration of the film recorder is also an important link in this chain of events. Digital film recorders can drift out of calibration over time and as you might suspect, this would be a bad thing if allowed to happen. At Rhythm & Hues, we check our film recorder weekly. We shoot a series of 34 grey images and then compare the graph of the resulting density readings against a standard graph. We also check the camera for focus, alignment, flaring, aspect ratio and the convergence of the 3 colors of the lasers.

Another link in the process where color can unintentionally change is the film lab. We previously discussed how we try to minimize variables in the printing of film by using

LAD standards. This process conforms the chemical print bath to a standard.

However, the lab's chemical negative bath is also subject to variation, although film labs are much more careful with this bath since while you can create more than one print from a single negative, the negative itself can only be developed once. Nevertheless, we run our own check on the film lab's negative bath for color consistency. We do this by sending the lab a strip of film negative called a *sensitometric strip* along with our regular negative. The sensitometric strip is cut from a roll of previously exposed, undeveloped camera negative. This camera negative has been purchased from Kodak who shoots 100' of a grey slate of 21 grey patches with the use of a sensitometer. A sensitometer consists of a standard light source (D5500), an optical step wedge (think of 21 neutral density filters in a row), a shutter, and a way to move the film. This creates a carefully controlled step wedge on the film with as little intervening hardware between the light source and the film as possible.

Theoretically, use of the sensitometer produces film that is exactly consistently exposed throughout the 100'. Kodak sends us the 100' in dry ice and we also keep it refrigerated. Then when we send our regular negative to the film lab, we cut off a 2' long section from this as our sensitometric strip and it is developed with the rest of the film. When we receive the developed film back from the lab, we then read at least the high and low end of the 21 steps of grey with a densitometer and see from this whether the lab negative bath is up to standard or too "hot" or "cold" in one of the 3 color layers. We are not so interested in the absolute density readings as we are in the differences between this week's reading and previous readings since consistency is the primary goal.

Last Steps

Once Rhythm & Hues has delivered its final film negative, you might think that the process is over. Well, it ain't necessarily so. There is another step that commonly occurs that affects color, and that is *color timing*. This is where final color adjustments are made by the film lab (although there has been some recent experimentation with digital color timing), usually in conjunction with the cinematographer and the director. Visual effects houses are not typically involved with this step.

As mentioned previously, even within a sequence, the lighting of the live-action can change somewhat from shot to shot. Color timing is a way of trying to achieve color, brightness, and contrast consistency across shot boundaries. So while a visual effects house may have done its duty to seamlessly integrate its elements into a particular shot, that shot may need to be adjusted to fit better with the previous or next shot. Some color timing does take place in the beginning in the creation of match clips, but it is normal for a final color timing session to take place after all shots are delivered.

Lastly, of course, the viewing environment of a public movie theatre presents the last set of variables with regards to the film color pipeline. Darkness levels, age of the projection bulb, and age of the print are the most common factors to think about. Once again, visual effects companies have little control over these variables.

Summary

I have presented the color management pipeline which Rhythm & Hues uses for mainstream film and video applications. There are certainly other areas of color management that are used in the entertainment industry, or may be used in the future. These areas include digital cameras, digital projection, digital television, and digital color timing. As these processes become more common, I am sure their particular color management issues will become more widely discussed.

Acknowledgements

I would like to thank Keith Goldfarb and Will McCown for their time and patience in explaining to me the finer details of Rhythm & Hues' color management pipeline, and for coming up with the pipeline in the first place.

Figures

The figures for this section are formatted as separate color plates. All figures are the property of Rhythm & Hues, and may not be used without permission outside of these notes for any purpose except as previously granted to ACM SIGGRAPH.

Figure 1: Various examples of how different types of contrast can direct the eye. The left image demonstrates texture contrast. The center image, contrasts of hue and depth-of-field. The right image primarily shows value contrast. (© 2002, Pauline Ts'o.)

Figure 2: This single image contains one color palette determined by a natural world experience - moonlit night in a kitchen, and simultaneously another palette determined by a cultural world - the happy cracker and box needed to look as they would in daylight for brand recognition purposes. The two color palettes were reconciled by justifying the daylight look of the product with a spotlight. Note that depth-of-field contrast was also used to direct the eye towards the product. (© 2001, Nabisco Brands, Inc.)

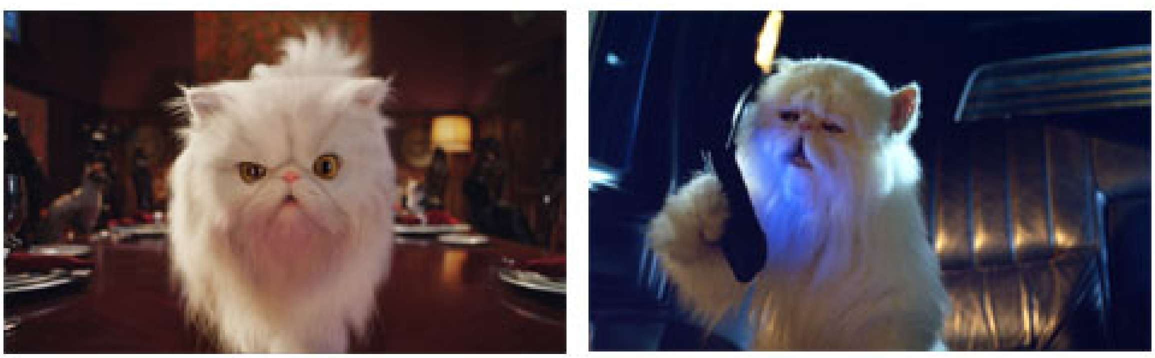


Figure 3: This completely computer-generated cat for “Cats & Dogs” needed to exactly match a live-action cat under different lighting situations. Sometimes shots would even alternate between the live-action cat and the cg cat, as was the case in the image on the left. Therefore, much care was taken with the color and lighting of every part of the cg cat. (© 2001, Warner Bros. Pictures.)

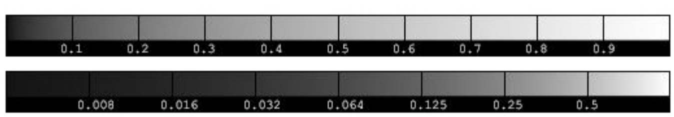


Figure 4: The top image shows a gradient which has the range 0.0 to 1.0, expressed linearly. The bottom image is a gradient with the range 0.004 to 1.0, expressed logarithmically. (© 2001, Rhythm and Hues, Inc.

Figure 5: A coarse 27 frame wedge of a computer generated polar bear. The wedge modified the red, green, and blue channels by +/- 1/2 stop. The center image (2nd row, 5th column) would be the original image without color modification. Note that increasing or decreasing all channels by a 1/2 stop is equivalent to increasing or decreasing the overall brightness by a 1/2 stop. (© 2000, Hankook Tire Co., Ltd.)

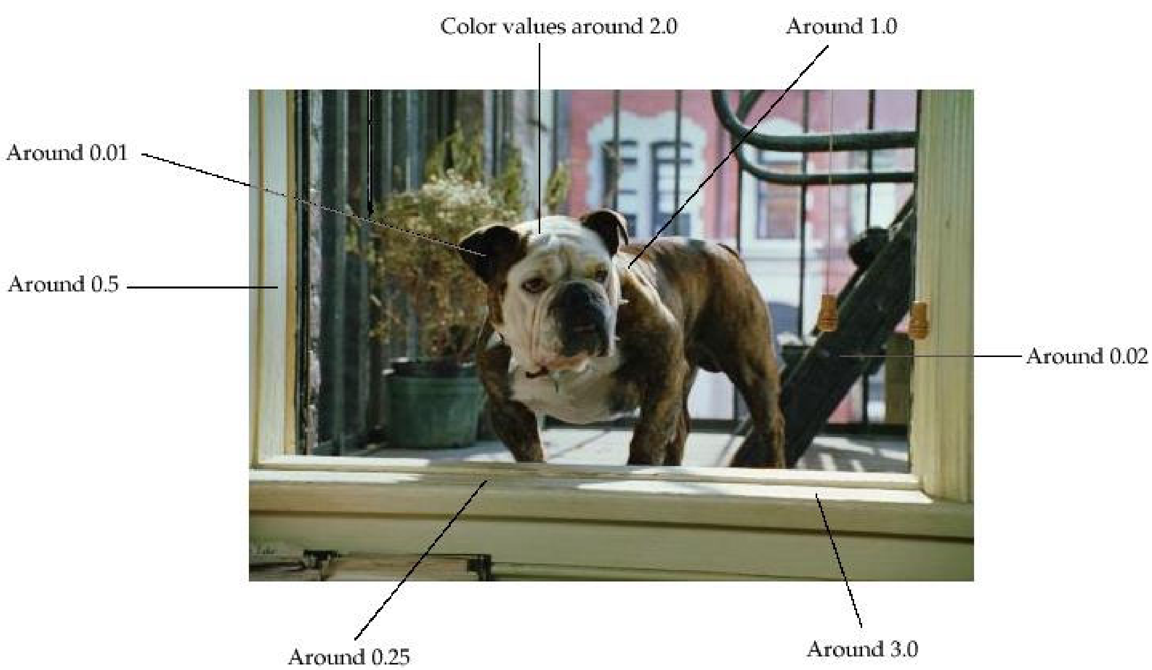
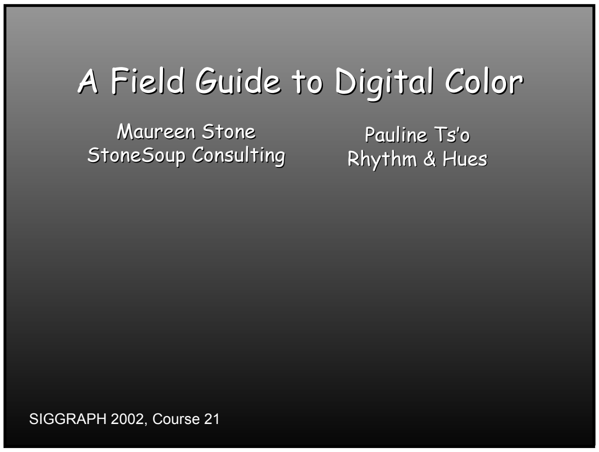
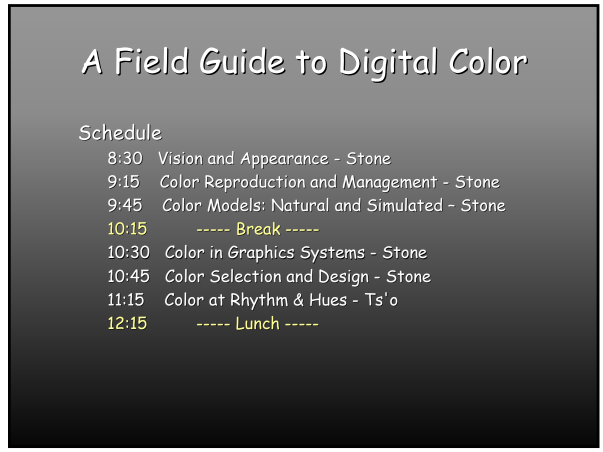


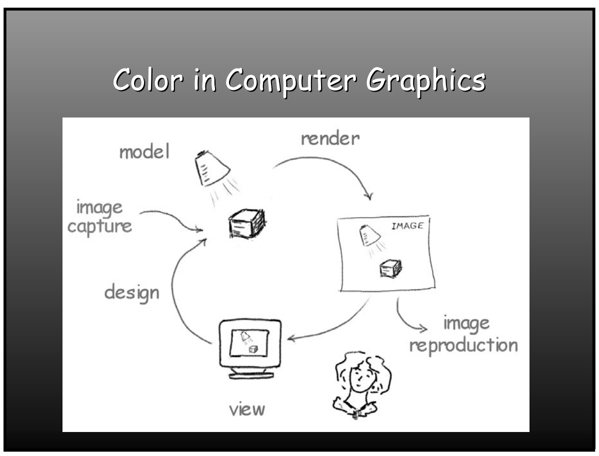
Figure 6: This image from "Little Nicky" gives a sense for what the values of the RGB triples in the rll file format mean. (© 2000, New Line Productions.)

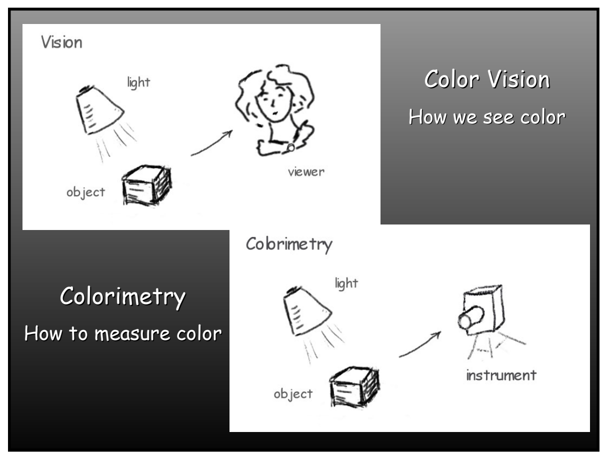
Figure 7: The right side shows an image and a zoomed up version of the same image as displayed in video mode. The left shows the same image and a zoomed up version as displayed in "quick film" mode. (© 2000, New Line Productions.)

Figure 8: In this shot from "Anna and the King" (with close-up on the right), the entire green/orange/gold roof, i.e., every part of the large building above the white wall, was a digital matte painting created in Photoshop. Color matching the gold tints and highlights as well as the overall contrast between the matte painting and the live-action wall was critical. (© 1999, Twentieth Century Fox.)


A Field Guide to Digital Color
Maureen Stone
StoneSoup Consulting
Pauline Ts'o
Rhythm & Hues
SIGGRAPH 2002, Course 21


A Field Guide to Digital Color
Schedule
8:30 Vision and Appearance - Stone
9:15 Color Reproduction and Management - Stone
9:45 Color Models: Natural and Simulated – Stone
10:15 ----- Break -----
10:30 Color in Graphics Systems - Stone
10:45 Color Selection and Design - Stone
11:15 Color at Rhythm & Hues - Ts'o
12:15 ----- Lunch -----


Color in Computer Graphics
model
render
image capture
IMAGE
design
image reproduction
view


Vision
light
viewer
object
Color Vision
How we see color
Colorimetry
light
instrument
object
Colorimetry
How to measure color

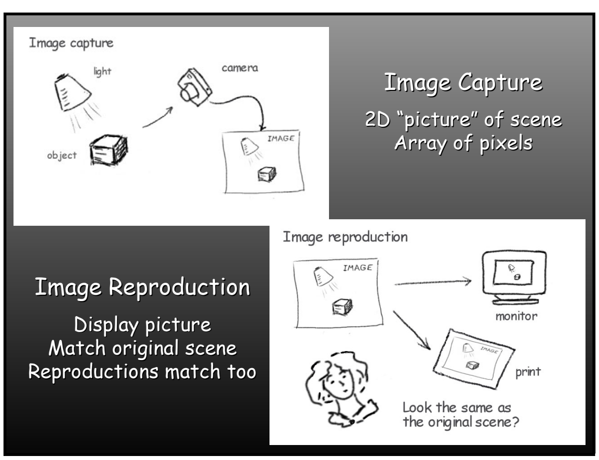
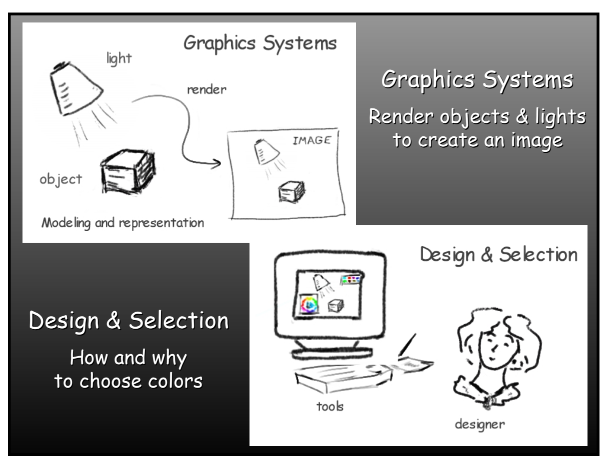
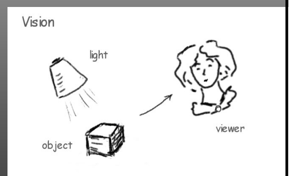


Image capture
light
camera
object
IMAGE
Image Capture
2D "picture" of scene
Array of pixels
Image reproduction
IMAGE
monitor
print
Look the same as
the original scene?
Image Reproduction
Display picture
Match original scene
Reproductions match too


Graphics Systems
light
render
IMAGE
object
Modeling and representation
Graphics Systems
Render objects & lights
to create an image
Design & Selection
tools
designer
Design & Selection
How and why
to choose colors

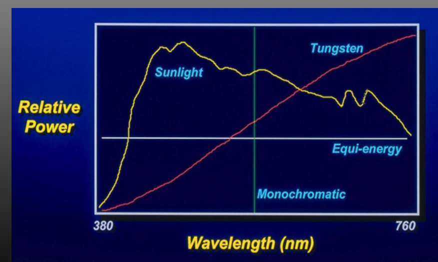
Color Vision

Light enters the eye
Absorbed by cones
Transmitted to brain
Interpreted to perceive color



Foundations of Vision
Brian Wandell

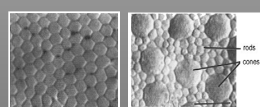
Spectral Distribution
Various Light Sources



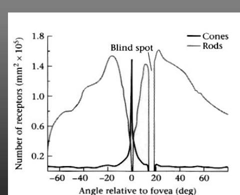
Cones and Rods

Cones: color vision
- Photopic vision
- Daylight levels
- Concentrated in fovea (center)

Rods: grayscale
- Low light levels
- Scotopic vision
- Peripheral vision



Fovial cones Cones and rods

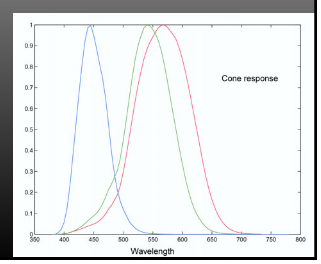


Distribution of rods and cones

From Foundations of Vision, fig 3.1 & 3.4,
© Brian Wandell, Stanford University

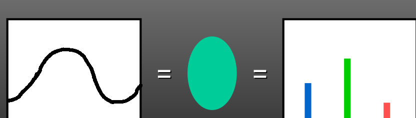
Cone Response Curves

Cone response is a function of wavelength
For a given spectrum
- Multiply by response curve
- Integrate to get response



Metamerism

Brain sees only cone response
Different spectra appear the same



Fundamental principle of
color reproduction

ColorWeb Demo

Java applets demonstrating color concepts
Brown University Graphics Group
- John Hughes, Anne Spalter
- Adam Dopplet, Jeff Beall, Robert George

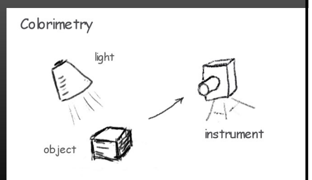
Electronic Schoolhouse Playground

Colorimetry

Measurement of color

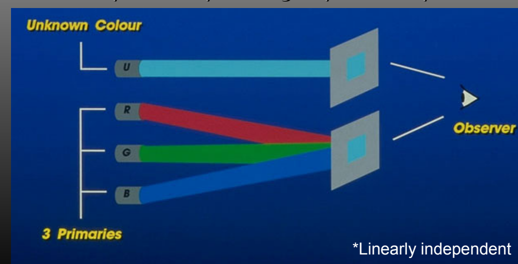
- Instruments that work like the eye
- Industrial and scientific application

Color Science
Wyszecki & Stiles



Color Matching Experiments

Match any color by adding any* three primaries



*Linearly independent

Tristimulus Values

Power of the primaries

- Define color by three numbers
- Specific to a set of primaries

Some restrictions apply

- Single colors
- Common viewing conditions
- Response, not appearance

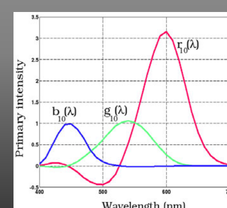
Color Matching Functions

To match any spectrum

- Multiply by CMF
- Integrate to get the tristimulus values

To create a set of CMF's

- Pick three primaries
- Match monochromatic colors
- Plot values vs. wavelength



R=645 nm, G=526 nm, B=444 nm
Styles & Burch

CIE Standard

Commission Internationale de l'Eclairage

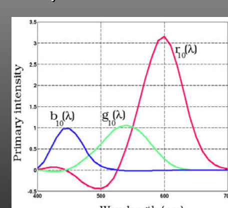
Formalized color matching

- Standard primaries
- Standard observer

Standard representation

- Linear transform
- CMF's all positive
- Easy to compute

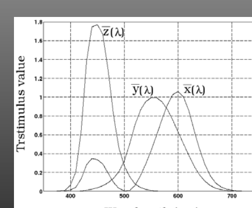
Styles and Burch



Measured basis

- Monochromatic lights
- Physical observations
- Negative lobes

CIE Standard



Transformed basis

- "Imaginary" lights
- All positive, unit area
- $\bar{y}(\lambda)$ is luminance

CIE Standard Notation

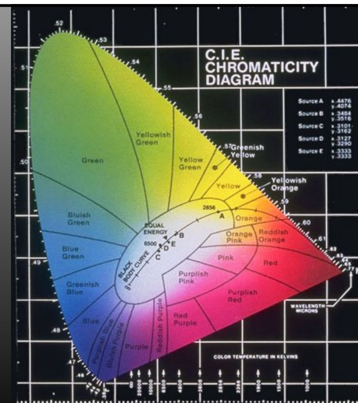
- Color matching functions
  - $\bar{x}(\lambda),\bar{y}(\lambda),\bar{z}(\lambda)$
  - Apply to measured spectrum
- CIE tristimulus values
  - X,Y,Z
  - Power of the CMF's
- CIE chromaticity coordinates
  - x,y or u, v
  - Standard 2D diagram

Chromaticity Coordinates

Project X,Y,Z on the (X+Y+Z)=1 plane

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

Colorfulness separate from Brightness



Courtesy of Photo Research

Color as a Linear System

- Grassman's additivity law
  - Spectrum: s = a + b
  - Response: $RGB_s = RGB_a + RGB_b$
- Linear basis
  - Color matching primaries
  - 3x3 matrix conversions
- Applies to tristimulus values and CMF's

Summary

- First level visual processing
  - Trichromacy—Three cones respond to spectrum
  - Metamerism—response, not spectrum
- Colorimetry
  - Color measurement
  - Color matching functions & tristimulus values
- But does what it LOOK like?

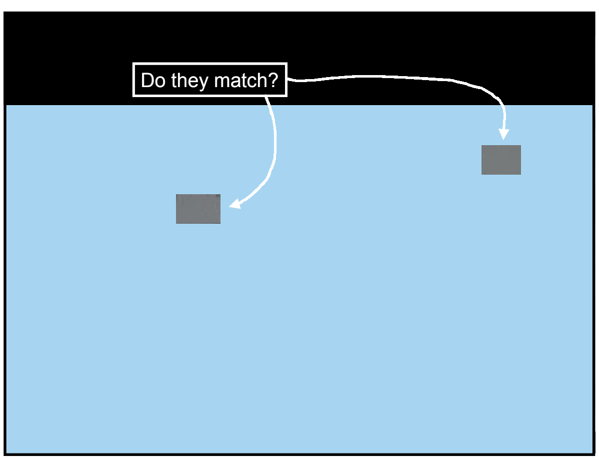
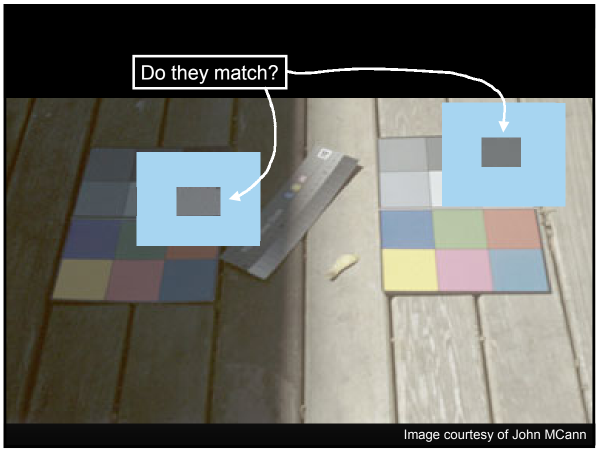
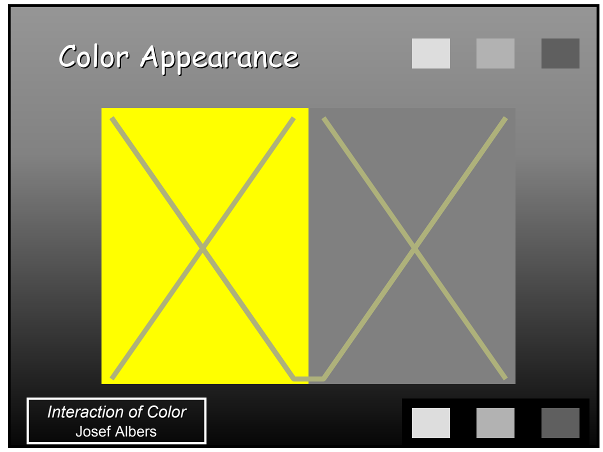
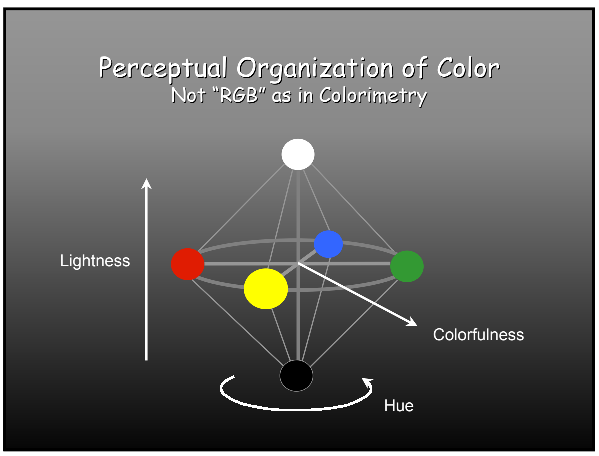
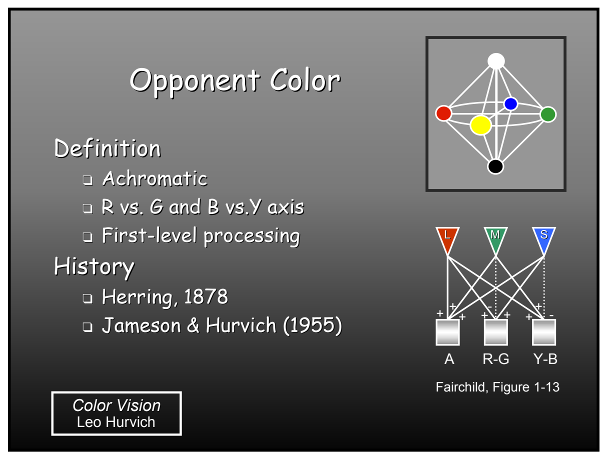


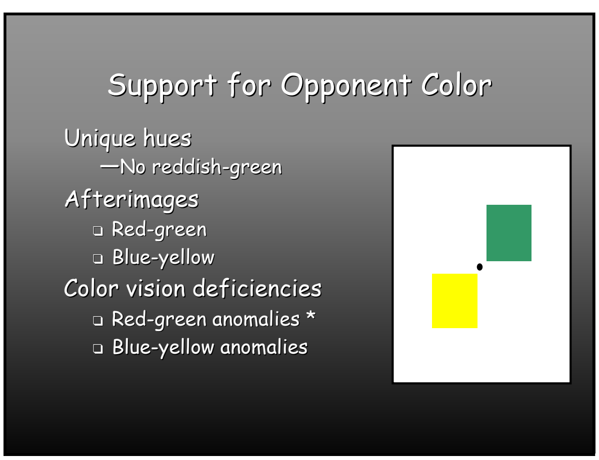
Image courtesy of John MCann

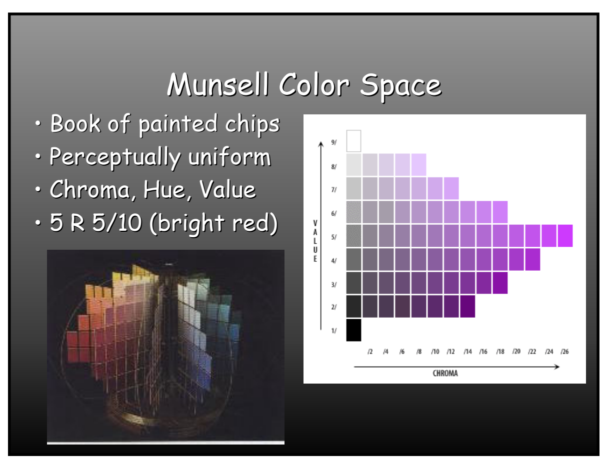

Do they match?
Image courtesy of John MCann


Color Appearance
Interaction of Color
Josef Albers


Perceptual Organization of Color
Not "RGB" as in Colorimetry
Lightness
Colorfulness
Hue

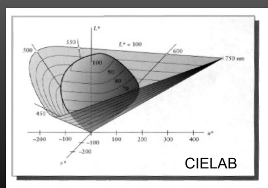

Opponent Color
Definition
Achromatic
R vs. G and B vs.Y axis
First-level processing
History
Herring, 1878
Jameson & Hurvich (1955)
Color Vision
Leo Hurvich
L
M
S
A
R-G
Y-B
Fairchild, Figure 1-13

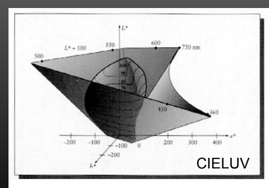

Support for Opponent Color
Unique hues
—No reddish-green
Afterimages
Red-green
Blue-yellow
Color vision deficiencies
Red-green anomalies *
Blue-yellow anomalies


Munsell Color Space
• Book of painted chips
• Perceptually uniform
• Chroma, Hue, Value
• 5 R 5/10 (bright red)
VALUE
CHROMA

CIELAB and CIELUV
Color Difference Spaces

1 unit = "just noticeable difference" (1 jnd)
Lightness (L*) plus two color axis
Function of XYZ of color plus XYZ of white





From Principles of Digital Image Synthesis by Andrew Glassner. SF: Morgan Kaufmann Publishers, Fig. 2.4 & 2.5, Page 63 & 64

Color Appearance Phenomena

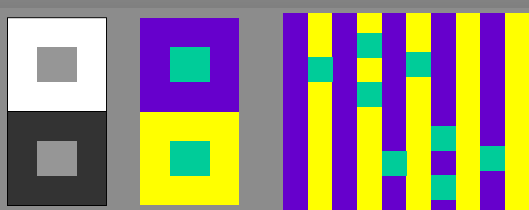
Related colors
- Simultaneous contrast
- Spatial frequency

Adaptation
- Light/dark adaptation
- Chromatic adaptation

Perceptual and cognitive effects

Color Appearance Models
Mark Fairchild

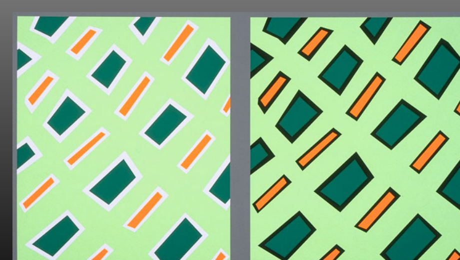
Simultaneous Contrast



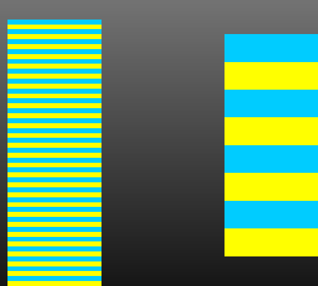
"After image" of surround adds to the color

Cognitive component
Identical patch and surround

Impact of Outlines
The Bezold Effect



Effect of Spatial Frequency
"Spreading" Effect



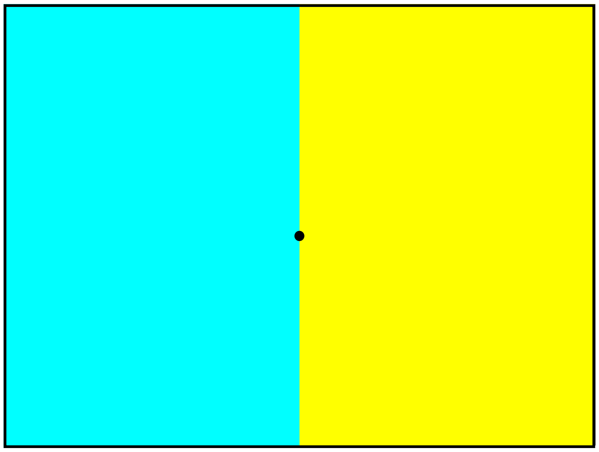
Redrawn from Foundations of Vision, fig 6

Adaptation

Light/Dark adaptation
- Uniform gain control
- Starlight to full sun

Chromatic adaptation
- Automatic "white balance"
- Scale cone sensitivities (von Kries)

Both physiological and cognitive
- Discount the illuminant
- "Object constancy"



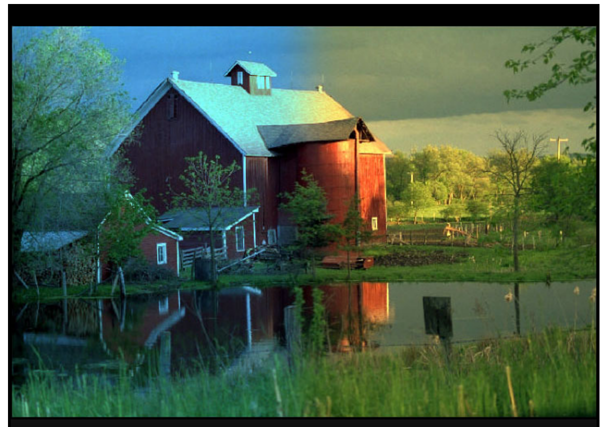


Image courtesy of Mark Fairchild

Original image | Overall Purple Tint | Tint Shirt Only

Inspired by Hunt's yellow cushion

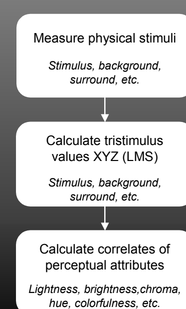
Color Appearance Models

From measurements to color appearance

Models

- CIELAB, RLAB, LLAB
- CIECAM97s
- Hunt
- Nayatani
- Guth, ATG

Color Appearance Models
Mark Fairchild



Color Vision Summary

First level visual processing

- Light on the retina (rods and cones)
- Trichromacy and metamerism
- Color measurement

Opponent color

- (R-G, B-Y, A)
- 1st level perceptual organization
- Perceptual color spaces

Color Vision Summary

Related colors

- Simultaneous contrast
- Spatial frequency

Adaptation

- Light/dark
- Chromatic (von Kries)

Appearance models

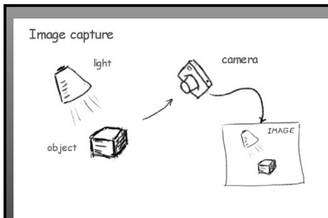
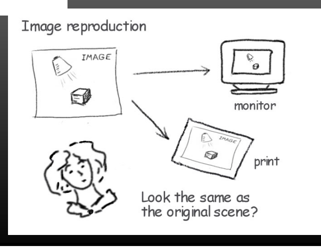


Image Capture

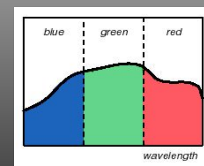
2D "picture" of scene
Array of pixels

Image Reproduction

Display picture
Match original scene
Reproductions match too



Color Reproduction



Principles
- Capture image as "RGB" components
- Reproduce each component
- Metameric match to the "original"

Understanding Digital Color
Phil Green

The Reproduction of Colour
Robert Hunt

Image Reproduction

Image capture
- Cameras
- Scanners

Different output media
- Additive: Video, displays, projectors
- Subtractive: Film, printing
- Media-specific characteristics and controls

Create image that "looks like" the original

Reproduce Image Appearance

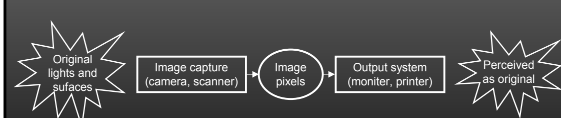
Media limited with respect to nature
- Device and image "gamuts"
- "Good reproduction" is application dependent

Tone reproduction
- Map the achromatic (brightness)
- Key to good reproduction

Color "correction"
- Simplest: make the primaries match
- Similar to RGB or CMYK balance adjustments

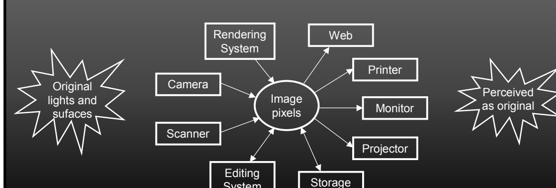
Traditional Image Reproduction

Media-specific pipeline
- Television, printing, photography
- Adjust any part of the process



Digital Color Reproduction

Multiple sources and outputs
- More dependent on standards
- Device-independent color management



Device-Independent Color

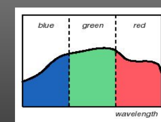
Characterize with respect to a standard
- Take measurements
- Map to standard space (CIELAB or CIEXYZ)

Device gamuts
- All reproducible colors
- 3D volume

Image gamuts
- Formal specification of image colors
- Transform to create a "good reproduction"

Image Capture

Create R, G, B color separations
- View through R, G, B filters
- Filters act as CMF's



Cameras
- CCD/CMOS array to capture natural scenes
- Focus, depth of field, exposure, etc.

Scanners
- Scan bar and controlled light
- Capture prints or transparencies

Image Characterization

RGB capture filters rarely colorimetric
- Mis-code image colors
- No general tranform to characterize image

Scanners
- Measure patches for specific medium
- Specific films, etc.

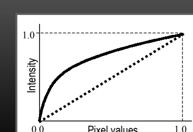
Cameras
- Characterize w.r.t. display
- General solution

Image Encoding

Encode R, G, B components
- Pixel = amount RGB
- Digital data will be quantized

Specify RGB with respect to standard

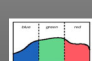
Specify pixel to intensity mapping
- Linear vs. nonlinear
- Non-linear is more efficient
- More later…

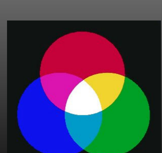


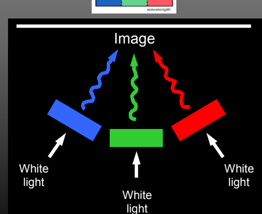
Additive Reproduction

Combine red, green, blue lights

Primaries
- red
- green
- blue



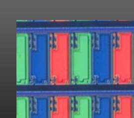




Displays

CRT Displays
- RGB phosphor dots and electron gun
- Intensity defines grayscale

LCD Displays
- Backlight and RGB filters
- Degree of transparency controls intensity



Digital Projectors

Imaging elements
- LCD: light valve
- DMD: pulse encoded

Three elements
- Dichroic mirrors split light
- Overlay to combine

One element
- Color wheel
- Combine sequentially

Television/Video

Camera captures separations
- Standard RGB
- Companded intensity
- Compressed to brightness + color (YCrCb)

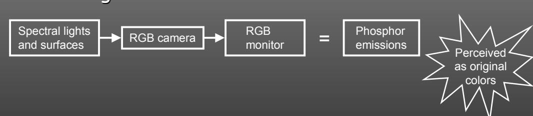
Reproduced on RGB display
- Decompress and map to RGB voltages
- Create correct appearance

Details are all specified

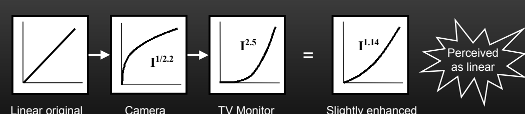
Digital Video
Charles Poynton

Television Pipeline

RGB Encoding



Intensity Encoding (ITF)



RGB Characterization

RGB is similar to tristimulus values

Easy to characterize
- Measure the primaries
- Measure the ITF's
- Linear RGB to XYZ is an invertible matrix

Monitors, LCD's: 3x3 matrix

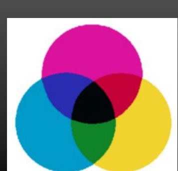
Projectors: 4x4 homogeneous matrix*

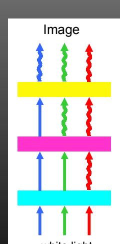
*Except for small DLP projectors

Subtractive Reproduction

Filter white light to modulate R, G, B

Primaries
- cyan
- magenta
- yellow

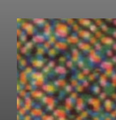


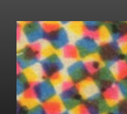


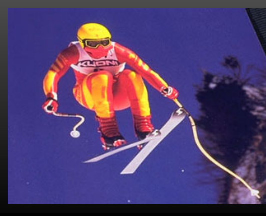
Subtractive Media

Photographic slides and prints
- CMY emulsion layers
- Density encodes grayscale

Ink on paper printing technologies
- CMY plus black (K) ink
- Halftone or dither patterns
- Some true grayscale
  - Inkjet multi-drop
  - Dye sublimation







CMYK Separations
red = magenta + yellow
green = yellow + cyan
blue = cyan + magenta
black for contrast

CMY Characterization

No simple transform to RGB or XYZ
- Inks are filters, not lights
- Color depends on light source

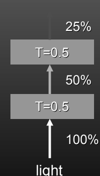
Characterize by sampling
- Establish standard setup
  - Lighting, ink, paper, processing, etc.
  - Gray balance, black printer
- Print and measure color patches (~1000)
- Interpolate 3D table: CMY to CIELAB
- Automated measurement systems

Density Measurements

Density = $\log_{10}(1/T)$
- T = normalized transmittence (or reflectance)
- D of 0 = clear, 3 to 5 = opaque
- T = 0.001 (0.1%), D = 3.0
- Product of T = sum of D

Densitometry
- Many kinds: optical, visual, color
- NOT colorimetry

Process control



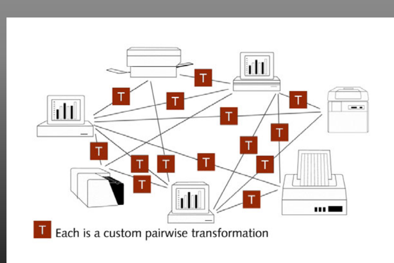
Digital Color Management

Device-independent color
- Based on color psychophysics
- CIE standards and extensions

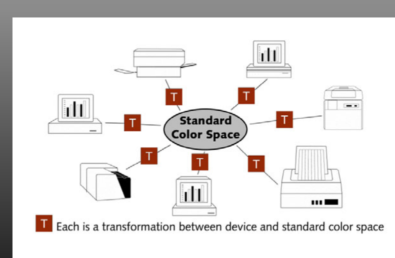
Color management systems
- Standard color specification (CIE)
- Characterize devices, images
- Transform colors/images to "look the same"

Digital Color Management
Giorgianni & Madden

The Problem



The Solution



CMS Components

Profile connection space (PCS)
- Usually CIELAB

Profiles
- Native color space to PCS
- ICC standard profiles (www.color.org)
- For each output device and image

Color management system (CMS)
- Implements transforms
- Different CMS can give different results

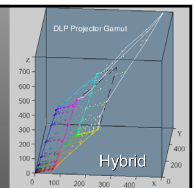
Gamuts and Gamut Mapping

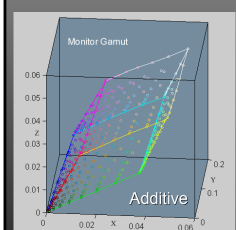
Gamut
- Set of all possible colors
- 3D volume in PCS, defined by profile

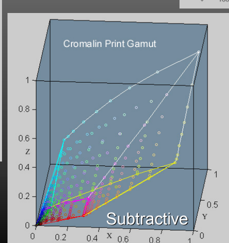
Gamut mapping
- Not all gamuts the same
- Must map out-of-gamut colors
- Follow the principles of image reproduction
- Application and image specific

Output Device Gamuts

Plotted in CIEXYZ







RGB color cube to XYZ

Chromaticity Representation

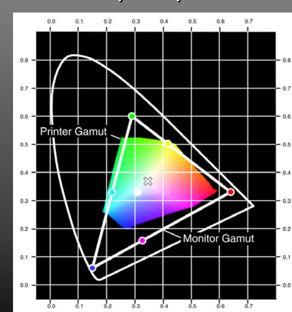
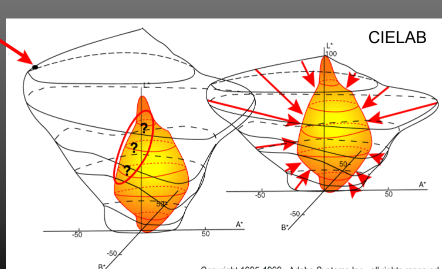


Image courtesy of ACM TOG

Gamut Mapping

Where does this color go?



Gamut Mapping Issues

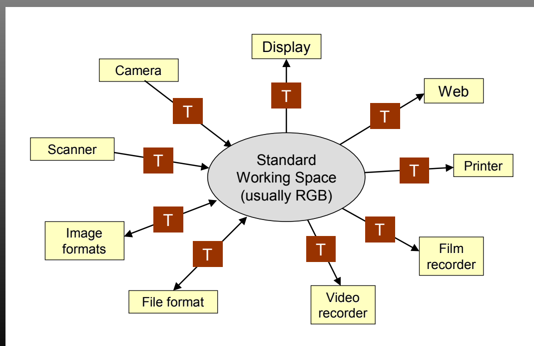
Device-to-device vs. image-to-device
- Generally device-to-device

Rendering intent
- Perceptual, relative colorimetric
- Saturated, absolute

Gamut mapping space
- Current systems use CIELAB
- Next generation: CIECAM 97 (?)

Extreme map is always unsatisfactory

User's View of Color Management



RGB Spaces

Fully specified RGB
- R,G,B, W chromaticities (CIE xy, color temp.)
- Max Y (luminance of white)
- Or, CIEXYZ for R, G, B
- Define with respect to PCS

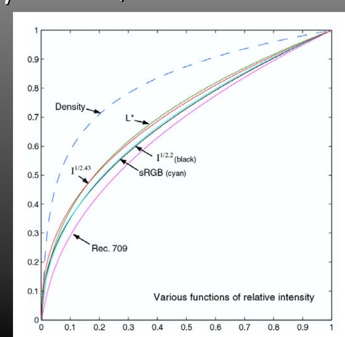
Linear or non-linear w.r.t. intensity
- Linear: graphics rendering, image processing
- Non-linear: displays, video and image formats
- Choice depends on application

Linear vs. Non-linear

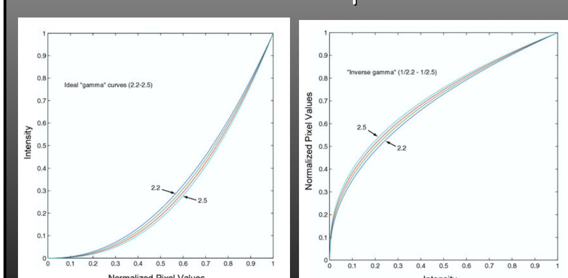
Linear intensity (or luminance)
- Physical measurement
- Not perceptually uniform

Non-linear
- Pixel = F(I), where I = relative intensity
- Form of F is
  - $k_1 I^{1/v}$ + offset for v around 2–3
  - $k_2 I$ for I < small value
- Creates (nearly) perceptually uniform scale

Many Useful, Non-linear Functions



Pixels/Perception



Control Transforms

Devices
- Standard settings
- Profiles
- Using default profiles

Standard formats
- sRGB
- JPEG, XIFF, PhotoYCC, video

Turn on color management

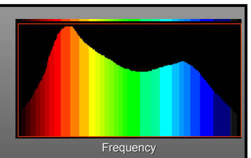
Image Reproduction Summary

Image capture: RGB to pixels

Image reproduction
- Additive (RGB)
- Subtractive (CMY+K)

Digital color management
- Common color space
- Know your transforms
- Look the same...as possible

Color in Nature



Colored light
- Spectral colors
- Refraction: Prisms, rainbows

Reflective
- Objects and lights
- Highlights, textures
- Subsurface scattering, fluorescence

Exploratory Color Playground

Color in Nature

Shimmering
- Thin films make interference patterns
- Butterflies, soap bubbles

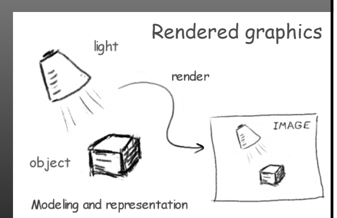
Scattered
- Fine particles scatter particular wavelengths
- Smoke, sky, eyes

The Color of Nature
Murphy and Doherty

Rendered Graphics

Virtual objects, lights and textures

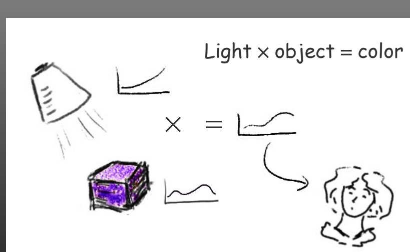
Rendered to produce a visible image



Illumination and Color in Computer Generated Imagery
Roy Hall

Principles of Digital Image Synthesis
Andrew Glassner

Objects & Lights



Spectral Color

Lights and object colors
- Sampled functions
- Wavelength vs intensity or reflectance

Convert to tristimulus values
- Color matching functions
- Integrate

XYZ to RGB for display

RGB from XYZ

Let X_R, Y_R, Z_R be the CIEXYZ values for monitor R, G, B, etc.

Then:

$$[R\ G\ B] \begin{bmatrix} X_R & Y_R & Z_R \\ X_G & Y_G & Z_G \\ X_B & Y_B & Z_B \end{bmatrix} = XYZ_{RGB}$$

Invert matrix to get RGB from XYZ
Works for any linear, additive RGB system

Physical Models

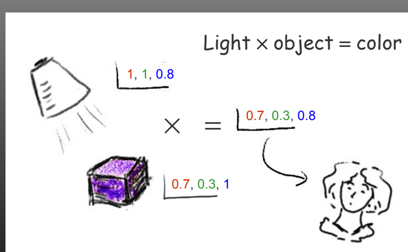
Advantages

- Realism
- Physical simulation
- Model more color phenomena

Disadvantage

- Complexity
- Expensive

Simplified Models



Based on RGB triples
Surface interactions also simplified

RGB Models

Simple extension of grayscale shading

- Red, green and blue intensity
- Works quite well for simple shading
- Inaccurate for complex models

RGB analogy

- Like color separations
- Like tristimulus values

RGB Accuracy

Accurately represents single color

Can't accurately represent interactions

- Lights with surfaces
- Reflections

Product of Tristimulus values	≠	Tristimulus values of Product of spectra

"Trichromatic Approximation..." Carlos Borges, SIGGRAPH '91

Color, Lighting, Shading

Color bound to lighting and shading model

Strauss model (OpenGL)

- Diffuse
- Highlight
- Ambient (shadows)

Renderman® shaders

Scene perception, not colorimetry

A Field Guide to Digital Color

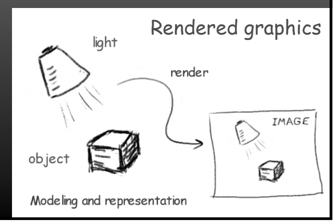
Schedule
- 8:30 Vision and Appearance - Stone
- 9:15 Color Reproduction and Management - Stone
- 9:45 Color Models: Natural and Simulated – Stone
- 10:15 ----- Break -----
- 10:30 Color in Graphics Systems - Stone
- 10:45 Color Selection and Design - Stone
- 11:15 Color at Rhythm & Hues - Ts'o
- 12:15 ----- Lunch -----

Color in Graphics Systems

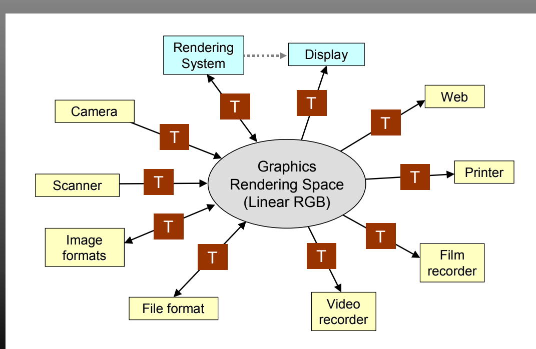
Display the Image

Rendering space
- Unbounded linear intensity
- Map to media gamut

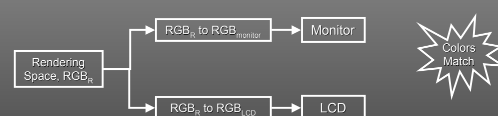
Mapping parameters
- RGB mapping
- Intensity mapping
- Desired appearance

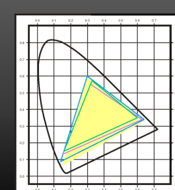


Graphics Working Space



RGB Mapping



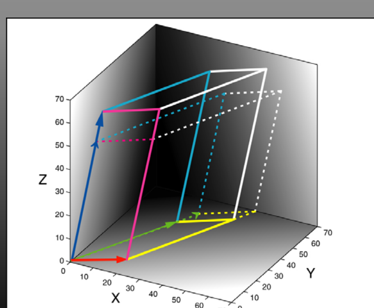


Data courtesy of Tom Fiske

$RGB_{monitor} \neq RGB_{LCD}$

Need 3x3 matrix for conversion

RGB in XYZ



RGB1 to RGB2

Assume R,G,B are linear in intensity

$$[R\ G\ B] \begin{bmatrix} X_R\ Y_R\ Z_R \\ X_G\ Y_G\ Z_G \\ X_B\ Y_B\ Z_B \end{bmatrix} = XYZ_{RGB}$$

where X_R, Y_R, Z_R are the CIE XYZ values for R, G, B

If
$RGB_1\ M_1 = XYZ_1$
$RGB_2\ M_2 = XYZ_2$

Then
$RGB_1\ M_1 M_2^{-1} = RGB_2$

Intensity Mapping

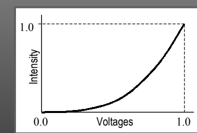
Linear rendering space

Non-linear display system
- Display gamma
- "Gamma correction"

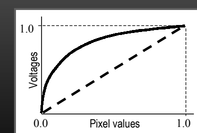
Desired appearance
- Bartleson & Breneman
- Tumblin and Rushmeier

The Gamma Problem

Device gamma
- Monitor: $I=A(k_1D+k_2V)^\gamma$
- Typical monitor $\gamma = 2.5$
- LCD: artificial monitor curve



OS Gamma
- Defined by operating system
- Inverse gamma curve ($I^{1/\gamma}$)
- "Gamma correction"



Display System Gamma

Product of Device and OS curves
- Divide Device by OS gamma
- $\gamma_{DS} = \gamma_D\ (1/\gamma_{OS})$

Display system gamma varies
- Different devices
- Different OS

Viewing conditions also affect perception of "gamma"

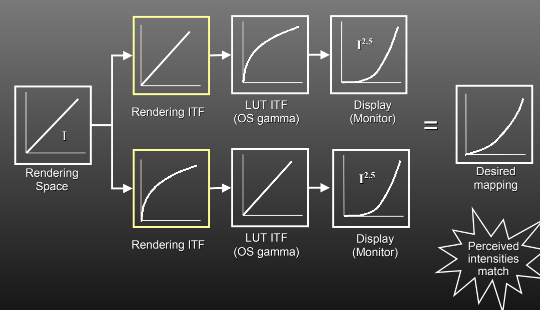
PC	Mac	SGI
1.0	1.4	1.7

Default OS Gamma

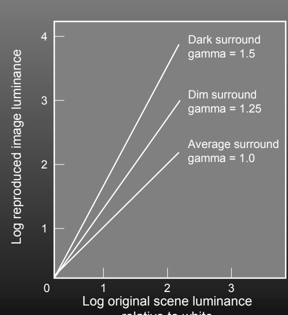
PC	Mac	SGI
2.2	1.6	1.3

Default DS Gamma

Intensity Mapping



Bartleson & Breneman



Increase contrast of reproduced image as function of the viewing environment.

Tumblin & Rushmeier

Tone reproduction operator based on human visual system

Considers
- Absolute brightness
- Adaptation
- Accurate display models

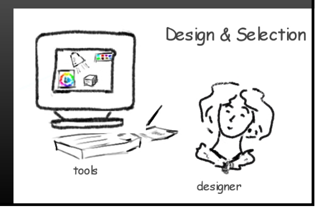
Resulting pictures more realistic

Graphics Systems Summary

- Simulate lights and objects
  - Representation
  - Create an image
- Display (reproduce) the image
  - Clearly define a working space
  - Lightness and RGB mapping
  - Gamut issues
  - Appearance issues

Design & Selection

- Visual designer's view
  - Color harmony, palettes
  - Visual function of the color
- UI designer's view
  - Interactive tools
  - Sliders and names

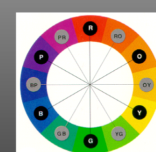


Principles of Color Design

- Selection of groups of colors
  - Visually appealing
  - Send a message
- Basic concepts
  - Color terminology
  - Color harmony
  - Color palettes

Principles of Color Design
Wucius Wong

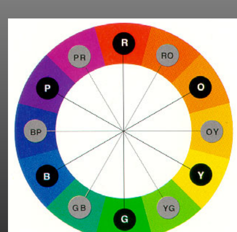
Color Terminology



- Hue
  - Color wheel
  - Primary, secondary, tertiary
- Chroma (saturation)
  - Intensity or purity
  - Distance from gray
- Value (lightness or brightness)

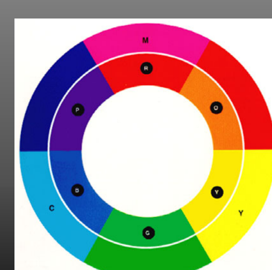
Color Wheel

- Primaries
  - Red, yellow, blue
- Secondaries
  - Orange, green, purple
- Complement
  - Color opposite
- Split complement
  - Two nearly opposite
  - Red, green-blue, yellow-green



Why Not CMY?

- Red, yellow, blue
  - Paint mixture
  - Came first
- Cyan, magenta, yellow
  - Printing inks
  - Transparent layers
- Similar mixing behavior
- Different gamuts



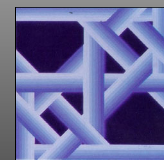
Tints and Tones

Tone or shade
- Hue + black
- Earth tones

Tint
- Hue + white
- Pastels

Color Harmony

Apply contrast and analogy to hue, chroma and value



Vary chroma

Analogous hues

Contrasting hues

Vary value

From *Principles of Color Design*, © 1997, John Wiley & Sons, Inc.

Palettes

Color sets
- 2-3 hues
- Tints and tones

Define mood or message
- Warm, soft, etc.
- Elegant, professional, etc.

Specific harmony
- Primary, analogous, etc.

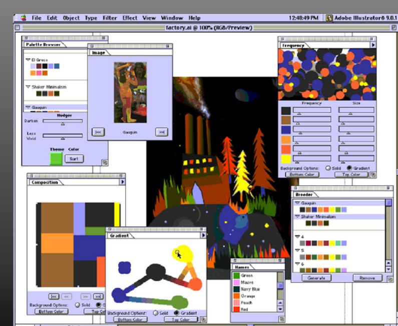
warm

elegant

analogous

primary

Interactive Palette Tools



Meier, Spalter, Karelitz, Simpson, Brown University

Color Selection Tools

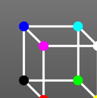
Select points in a color space
- Device-specific spaces
- Perceptual spaces
- Sliders and tools of many forms

Color names

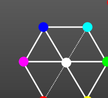
Little support for design principles

"Intuitive" Views of RGB

Transforms of the RGB color cube

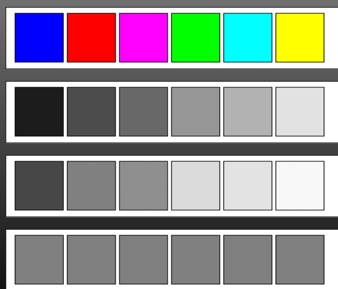


HSV
- Hue hexagon (piecewise linear)
- Add black, add white
- Tints and tones



HSL
- H and S as in HSV
- Lightness more like perceptual spaces

Perceived Lightness/Brightness

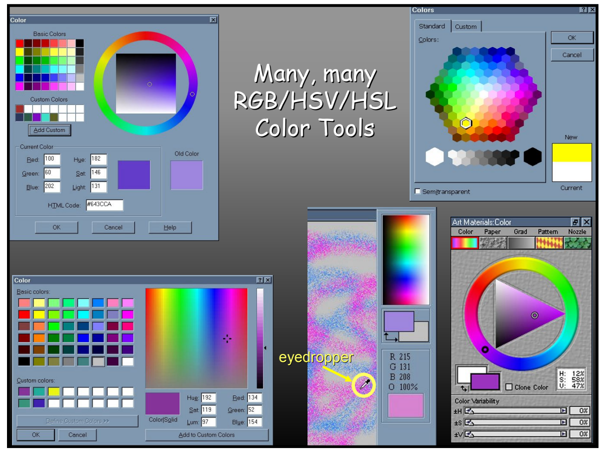


Corners of the RGB color cube

Luminance

L*

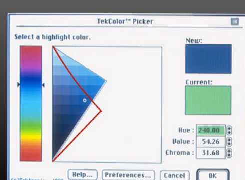
L from HLS



Perceptual Color Spaces

Perceptually uniform spaces
- Munsell, Ostwald
- CIELAB, CIELUV

Digital examples
- TekHVC/TekColor
- Metapalette
- Canon Color Advisor



Color Names

Perceptual/linguistic fundamental
- Berlin & Kay, 1969
- Johan Lammens, SUNY 1994

Commercially used
- *Color, Universal Language and Dictionary of Names*, Kelly & Judd, NBS 1976
- Berk, Brownston & Kaufman, 1982
- Pantone Matching System®

black
white
gray

red
green
blue
yellow

orange
purple
brown
pink

Other Applications

HTML names
- Named RGB values
- 140 names (!)

X-Windows
- Color name database
- Binds name to value
- /usr/lib/X11/rgb.txt

Color	Hex #	Color	Hex #
Aqua	00ffff	Navy	00005d
Black	000000	Olive	5d5d00
Blue	0000ff	Purple	5d005d
Fuschia	ff00ff	Red	ff0000
Gray	5d5d5d	Silver	a8a8a8
Green	005d00	Teal	005d5d
Lime	00ff00	White	ffffff
Maroon	5d0000	Yellow	ffff00

Information Visualization

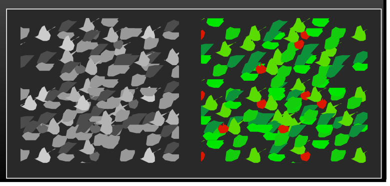
Above all, do no harm *—E. R. Tufte*

Color must be
- Clarifying, not confusing
- Tasteful
- Robust
- Friendly to all

Envisioning Information
Edward R. Tufte

Fundamental Uses: To Label

Color as a noun
- Identify (uniqueness and distinctness)
- Grouping and segmentation
- Layering (foreground and background)
- Highlighting and emphasis

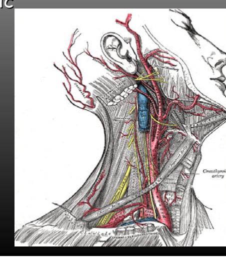


Information Vizualization
Colin Ware

To Represent or Imitate Reality

Color as representation
- Key color to real world
- Iconographic vs. photographic

Examples
- Blue for water
- Shading for shadows
- Gray's Anatomy

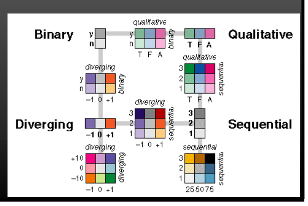


http://www.bartleby.com/107/illus520.html

Fundamental Uses: To Measure

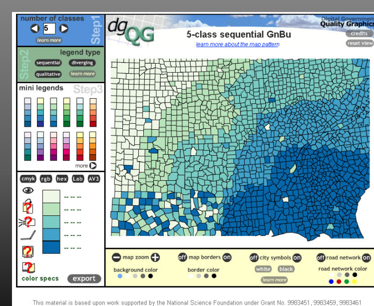
Color as quantity
- Few intuitive orderings
- Grayscale, saturation scales
- Creates clusters of similar values

Examples
- Thermometer scales
- Density plots
- Thematic maps



"Color Brewer" —Cynthia Brewer

Color Brewer



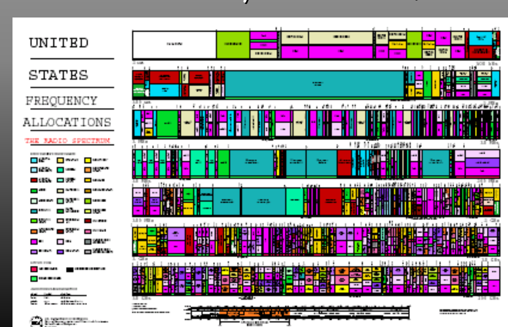
http://www.personal.psu.edu/faculty/c/a/cab38/ColorBrewerBeta.html

To Enliven or Decorate

Color as beauty
- Basic human desire
- Attractive
- Often ignored or belittled
- Best visualizations are beautifully designed

Examples
- Illuminated manuscripts
- Colored computer "desktops"
- Aesthetics are personal

Above all, do no harm



Making Color Effective

Avoid clutter
Emphasize legibility
Accommodate
- Atypical color vision
- Color-deficient devices
- Cross-media rendering

"Using Color Effectively in Computer Graphics"
Lindsay MacDonald, *CG&A*, July 1999

Avoid Clutter

Pre-attentive label
- Count the 7's
- Too many colors spoils effect

Examples
- Maps and charts
- Colored text, logos

674912497 54625
496087657 23514
650678726 14332
647612561 27810
467234512 15346
607237614 28385

Legibility

Luminance contrast
- Hue alone doesn't make an edge
- Make legible in grayscale
- Dropshadows and similar tricks

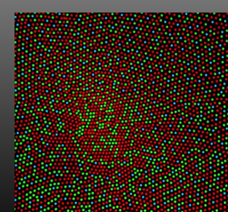
legibility	legibility	legibility	legibility
legibility	legibility	legibility	legibility
legibility	legibility	legibility	legibility
legibility	legibility	legibility	legibility
legibility	legibility	legibility	legibility
legibility	legibility	legibility	legibility

With drop shadows

The "Blue" Problem

"Blue" cones are sparsely distributed
Focus different than red and green cones
No blue cones in the center

Blue text can be
Hard to read
Different Focus
Different Focus
Different Focus
Different Focus


Fairchild, Figure 1-5

Accommodate the Viewer

Atypical color vision
- Minimize red-green encoding
- Duplicate all information
- Should "read well" in grayscale

Color-Deficient Media

Monochrome media
- Luminance contrast
- Duplicate information

Dither and halftone patterns
- Texture and raggedness
- Use solid colors
- Default colormaps

Netscape Navigator

Design & Selection Summary

Many clever tools
- Selecting individual colors
- Generally device-specific

Missing
- Full support for design principles
- Palette selection and management
- Levels of indirection, functional colors

Finale