Edward J. Giorgianni

Thomas E. Madden

DIGITAL COLOR MANAGEMENT

Encoding Solutions

Second Edition

WILEY

WILEY-IS&T SERIES IN IMAGING SCIENCE AND TECHNOLOGY

Digital Color Management

Digital Color Management: Encoding Solutions, Second Edition Edward J. Giorgianni and Thomas E. Madden
 ISBN: 978-0-470-51244-9

Wiley-IS&T Series in Imaging Science and Technology

Series Editor:
Michael A. Kriss

Consultant Editors:
Anthony C. Lowe
Lindsay W. MacDonald
Yoichi Miyake

Reproduction of Colour (6th Edition)
R. W. G. Hunt

Color Appearance Models (2nd Edition)
Mark D. Fairchild

Colorimetry: Fundamentals and Applications
Noburu Ohta and Alan R. Robertson

Color Constancy
Marc Ebner

Color Gamut Mapping
Ján Morovič

Panoramic Imaging: Sensor-Line Cameras and Laser Range-Finders
Fay Huang, Reinhard Klette and Karsten Scheibe

Digital Color Management (2nd Edition)
Edward J. Giorgianni and Thomas E. Madden

Published in Association with the Society for Imaging Science and Technology

IS&T
imaging.org

Digital Color Management

ENCODING SOLUTIONS

Second Edition

Edward J. Giorgianni
Senior Research Fellow (retired),
Eastman Kodak Company, USA

Thomas E. Madden
Senior Principal Scientist,
Eastman Kodak Company, USA

WILEY
A John Wiley and Sons, Ltd, Publication

This edition first published 2008

Registered office
John Wiley & Sons Ltd, The Atrium, Southern Gate, Chichester, West Sussex, PO19 8SQ, United Kingdom

For details of our global editorial offices, for customer services and for information about how to apply for permission to reuse the copyright material in this book please see our website at www.wiley.com.

Library of Congress Cataloging-in-Publication Data

Giorgianni, Edward J., 1944–
Digital color management : encoding solutions / by Edward J. Giorgianni & Thomas E. Madden. – 2nd ed.
p. cm.
Includes bibliographical references and index.
ISBN 978-0-470-51244-9 (cloth)
1. Image processing–Digital techniques. 2. Color. 3. Coding theory. I. Madden, Thomas E. II. Title.
TA1637.G56 2008
621.36′7–dc22

2008041558

A catalogue record for this book is available from the British Library.

ISBN 978-0-470-51244-9

Set in 10/12pt Times by Aptara Inc., New Delhi, India.
Printed in Italy by Printer Trento

Contents

Series Preface

The other day I used my new 12 mega-pixel Digital SLR camera to take some pictures of my wife's garden, which was in full bloom with Fall colors. The camera has two settings for color encoding and two rendering settings, normal and vivid. I used vivid. The file format was "raw", using 16 bits per pixel. I transferred the image, picking one of two possible renderings, to my computer using the provided camera software and converted the raw format to a TIFF format so I could use my favorite image-manipulation software package. I used some of the color, tonescale and sharpness features to make a final image that looked great on my LCD display. I then printed it on my inkjet printer, again using the software provided by the printer manufacturer, making sure that I picked the correct settings for paper type (glossy photographic paper) and a color calibration compatible with my LCD monitor. I then proudly showed the print to my wife, and to my surprise she was not impressed with my hard work. She said, "The reds are not as bright or saturated; and look at the blues, they are not right. Can't you do something to make them look right?" So I went back to the computer and used saturation and curve shaping tools to make the print look more like the "actual" scene in bright sunlight. When finished, the lavender flowers were almost an exact visual match (in bright sunlight) while the reds, blues, yellows and oranges were close enough to satisfy my wife. Greens were not important to her, but then she always wants me to remove grass for more flowerbeds (and weeds).

So why was the above process so difficult? Each of the imaging-system components (camera, monitor, software, and printer) was from a different manufacturer and each had its own color reproduction characteristics. In an "open system" this should not make any difference; but those who make their own prints know that if you change even the paper, let alone the printer, a "different looking" print will be obtained. So how does one approach this multifaceted problem of digital color management? The seventh offering of the Wiley-IS&T Series in Imaging Science and Technology, ***Digital Color Management: Encoding Solutions, Second Edition*** by Edward J. Giorgianni and Thomas E. Madden addresses these issues in a well thought out, systematic manner. This text is an absolute must for anyone working on color system management in the film, display, digital camera, printer, scanner, television and digital cinema industries, for it provides a clear discussion of the realities and myths of imaging system color reproduction along with a lucid introduction to the fundamentals of color vision and color reproduction.

To put the concept of "closed systems" and "open systems" in perspective consider the following examples. When I first joined Eastman Kodak in 1969, many of the researchers there were avid photographers and would love to argue about which color slide film was best. Many loved the Kodachrome blue skies, while others felt Ektachrome gave more realistic colors overall and some loved the reds of Agfachrome. Slide films represent the ultimate in closed systems in that the entire process is designed and (to a large extent) controlled by a single manufacturer. The spectral sensitivities were selected for either daylight or for tungsten scene illumination, and the dye sets were designed to give good skintone-to-neutral balance when combined with a well-controlled chemical process. The final judgment was made in a special projection room under dark-adapted conditions and with really "sharp," flare-free projectors that used the display illuminant specified for the film. Many "tweaks" were made to the emulsions and process chemistry until everything was just right (according to the "in-house experts"). This went on in every film manufacturer's R&D organizations trying to get the best professional or amateur film that their patentable technology would allow. The same was true in the color negative systems, where film spectral sensitivities (daylight and tungsten), colored couplers, image dyes, interimage chemistry (for sharpness, color correction and low grain), and matching photographic paper characteristics were designed to have good skintone-to-neutral balance, good color saturation and sharpness when processed properly. Different formulations were used for amateur and professional systems to meet different customer needs, but the goal was always a good picture with little effort on the customer's part. Some film systems, such as motion picture films, were designed to have greater flexibility in order to provide greater artistic expression. However, in all cases, the systems were closed and often proprietary.

Today, to a great extent, "open-systems" have pushed much of the responsibility for a good picture to the user rather than the provider. There are countless combinations of films or digital cameras, scanners, printers (inkjet, color electro-photographic, thermal dye-transfer, etc.) and processing software that can be used by consumers and professionals alike. These choices are often up to the user, and there is no way to ensure that each combination will give the same image or even one that the user will like without some sort of hands-on manipulation. This dilemma is acute for printing services (or companies that make digital scanners and printers) that receive images originating on film or from digital files from customers who expect good quality prints with realistic, pleasing colors. Given all the variability in the system, how can this be done? The second edition of ***Digital Color Management: Encoding Solutions*** provides the most comprehensive and systematic approach to solving these issues. The authors have done a masterful job in defining the issues and problems, showing systematic solutions and pointing out that one "color-management methodology" does not fit all color-imaging systems. This single offering cannot provide explicit solutions to all color-management problems; but it does provide serious color scientists and engineers the basis from which they can formulate solutions to their particular color-management problems.

Edward J. Giorgianni is a product of his formative days (early 1970s) in the Eastman Kodak Research Laboratories. Ed and I worked in the Color Physics group in the Color Photography Division. At that time our responsibilities focused more on image structure than color, but we did some self-education in color using Dr. Robert Hunt's text on ***The Reproduction of Color, Second Edition***. Ed, already intrigued with color, decided to make the study of color his career goal. He succeeded beyond his own expectations, making many valuable contributions within Kodak and outside with his publications, and his and Thomas E. Madden's first edition of ***Digital Color Management*** in 1998. Ed took full advantage of learning from some of the giants of the world of color science including Frank Clapper, Jim Bartleson, Ed Brenemen, LeRoy DeMarsh, Ralph Evans, Bob Hunt, Dave MacAdam and Daan Zwick. He brought to his efforts a systematic approach that was understood in principle by his colleagues, but which had not previously been codified in a formal way. Ed retired in 2005 as a Senior Research Fellow from Eastman Kodak Company, having worked for thirty-eight years designing advanced photographic and electronic color-imaging products and systems. He is currently an adjunct instructor at the Center for Imaging Science at the Rochester Institute of Technology and an independent consultant to several corporations and professional groups, including the Academy of Motion Picture Art and Sciences. A Kodak Distinguished Inventor, he holds more than thirty patents in the fields of imaging technology and digital color management. He has taught courses in color science and color imaging for many years and is a four-time recipient of the Kodak Imaging Science and Technology Instructor of the Year Award. He is the author of numerous technical papers and a contributing author to four textbooks on color imaging and color management.

Thomas Madden, a Senior Principal Scientist at Eastman Kodak Company, received his BS and MS degrees at Valparaiso University and the Rochester Institute of Technology, respectively. He joined Kodak in the early 1980s and attended in-house color courses taught by Ed Giorgianni while working on applications mating digitized photographic images and electronic displays. The student-teacher relationship grew into a strong collaboration over the next 20 years where they worked together on a series of digital imaging projects including the Kodak Premier Image Enhancement System and the Kodak Photo CD System. Their collaborative work on these projects and continuous discussions about color imaging systems led to their Unified Paradigm that is the basis for their color management concepts as codified in ***Digital Color Management, Second Edition.*** Tom took over many of Ed's teaching responsibilities upon Ed's retirement, thus continuing the tradition of in-house color education at Kodak, where he also received the Instructor of the Year Award. Tom designs digital color-imaging products for consumer and professional applications, holds numerous patents, and continues to publish, lecture and teach about color management around the world.

The offering is a result of their efforts to formalize their collective thoughts and knowledge of color management.

MICHAEL A. KRISS

Formerly of the Eastman
Kodak Research Laboratories
and the University of Rochester

Acknowledgments

We wish to thank the following technical reviewers of this second edition for their many helpful contributions, comments and suggestions: Paula J. Alessi, Gabriel Fielding, Alexander G. Forsythe, Joseph Goldstone, Thomas O. Maier, Brian E. Mittelstaedt and Richard G. Patterson.

Thanks also to Carolyn J. Mauro and Neville Hankins for their copyediting of our manuscript and to the entire staff at John Wiley and Sons, Ltd., especially Beth Dufour, Alexandra King, Nicky Skinner and Simone Taylor for their efforts in the production of this book. We also wish to thank David B. Goldstein for providing many of the included photographs, Frank R. Brockler for preparing the three-dimensional figures and Michael A. Kriss for providing the preface.

Introduction

Not very long ago, digital color technology was available only on high-end color-imaging systems. Its use essentially was limited to commercial applications such as the production of graphic arts prints and motion picture visual effects. Although high-end systems remain an important segment of the digital color-imaging industry, the overall field has changed significantly. Inexpensive high-quality digital cameras, scanners, computers, electronic displays, and color hardcopy output devices such as inkjet, laser, and thermal-transfer printers are readily available. As a result, digital color imaging is now pervasive in virtually all scientific, commercial, and personal applications.

These changes have profoundly affected every aspect of digital imaging. In particular, they have made it necessary to fundamentally alter the way images are *color encoded*, i.e., how the colors that make up images are numerically represented in digital form.

In the past, the color-encoding methods used on high-end systems typically were quite straightforward. That was possible because these systems generally were self-contained or "closed." For example, some electronic prepress systems always used a certain type of scanner to digitize images input from a particular type of photographic film, and the resulting digital images were used exclusively for graphic arts printing. Relatively simple color-encoding methods were successful in such systems due to the invariant nature of the input and output devices and media.

All of this has changed and continues to evolve as more and more digital color images are being captured or generated for various types of computer displays and related applications, for presentation on conventional and high-definition televisions, and for output to an ever-increasing variety of hardcopy devices and media. It has become essential, then, to move from "closed" systems to systems that are "open."

By definition, an "open" system is not restricted to using only certain inputs and outputs; it can make use of all available types of imaging devices and media. Also implied by definition is that if every digital color-imaging system were "open," digital image data could be interchanged freely among all systems, with predictable color results.

When the subject of such "open" systems is raised, the discussion invariably gravitates toward a concept widely referred to as *device-independent color*. "Device-independent" in this context means that color is expressed and exchanged in a way that does not depend on the particular characteristics of any given imaging device or medium.

Virtually all proposals for such "device-independent" color have been, and continue to be, based on standard color-measurement techniques. However, our experience is that color-encoding methods based on these techniques alone are not sufficient for creating truly open systems capable of supporting disparate types of devices and media, nor can they alone provide for the unrestricted interchange of images among different types of imaging systems.

This is not an obvious or widely accepted position to take. After all, standard methods of color measurement have been used successfully for decades for any number of applications. It would seem reasonable that the use of these well-established measurement methods, along with some well-defined file format standards, should allow the open interchange of digital color images among systems. Nevertheless, despite all the attention such color-encoding methods continue to receive, actual experience has shown that they work only in the most restricted applications.

This does not mean, however, that it is impossible to design systems that are truly open. Nor does it mean that color measurement would not be an important part of such systems. What it *does* mean is that to function properly, the digital color encoding of open systems must be based on advanced color-encoding methods that go well beyond standard color-measurement techniques.

A principal objective in this book is to describe these advanced color-encoding methods along with a number of important system-specific methods for digitally representing color. Our broader objective is to cover all the information required for a solid understanding of the technology of representing and managing color in the digital domain. It is impossible, however, for a single book of reasonable size to include all details regarding every issue that could be discussed. Many implementation-related topics, such as color metrology, color-appearance modeling, 3-D table construction and interpolation, etc., are sufficiently complex to require entire books devoted to those single topics. Because such references are available elsewhere, we will allot less coverage here to explaining precisely *how* certain color operations might be implemented and instead concentrate on fully explaining exactly *what* those operations are intended to accomplish and *why* they are necessary for successful color management.

We recognize that many who require this information may not have backgrounds in color science or color imaging. For that reason, we begin with a part entitled *Fundamentals*, which deals with some basic principles of color measurement and color imaging. We have tried to make this part as concise as possible by covering only the information necessary for understanding digital color encoding. (As a result, it may well be the only discussion of color science ever written that does *not* include the seemingly mandatory cross-sectional diagram of the human eye!)

The second part, *The Nature of Color Images*, describes the color properties of color images produced on various types of imaging devices and media. Much of this information is not widely available, which has been unfortunate. We think the lack of factual information on this subject has been responsible for many common misconceptions regarding color imaging in general and color encoding in particular. Again, the discussion in this part is aimed toward our principal subject. Our objective is to explain why images from various types of devices and media *must* differ fundamentally in their basic color properties. These differences must be understood in order to appreciate the problems of encoding and managing color and to understand the solutions to those problems.

The final two parts cover our main topics: digital color encoding and the color-managed systems in which that encoding is used. These parts include the following:

- An explanation of two fundamentally different methods of representing color images: scene-based color encoding and rendered-image-based color encoding.
- A discussion of various popular myths and misconceptions regarding device-independent color and other related topics.
- A discussion of three basic color-management paradigms that describe the different behaviors of various types of color-imaging systems.
- A description of a unified paradigm for color management—a concept that offers the promise of an integrated color-managed environment for the color-imaging industry.
- A description of a practical, appearance-based color-encoding method capable of supporting the Unified Paradigm and of providing unrestricted communication of color within and among all digital color-imaging systems.
- A series of examples demonstrating how color management based on the Unified Paradigm can be applied in a wide variety of color-imaging situations.
- Detailed descriptions of two complex color-managed systems based on the Unified Paradigm, one of which supports a highly disparate array of inputs and outputs, and the other designed to meet the specific needs of digital cinema.

The book also contains a comprehensive glossary and an extensive series of appendixes that provide additional information on selected subjects and detailed descriptions of various color calculations and transformations.

We suggest this book be read straight through, even if the reader is familiar with most of the background material presented in the earlier sections. Each chapter builds on the preceding discussions, and most contain information that is not commonly available. If a chapter must be skipped, we would urge that at least the Summary of Key Issues given at the chapter's end be read.

PART I

Fundamentals

In this introductory part, some basic principles of color and color measurement will be examined in the context of color-imaging systems. This examination will provide the foundation required for later discussions on color images, color encoding, and color management.

The part begins with a review of the techniques of color measurement, which are the bases of all methods of numerically representing color. Color-imaging systems then will be described—not in terms of specific technology, but in terms of the basic functions they must perform. The focus here, and throughout the book, will be on systems for which the ultimate goal is to produce images that are high-quality color reproductions of original images.

Two very different types of original images will be dealt with in these discussions. In some cases, the original will be a live image, such as an outdoor scene being recorded with a digital still camera. In other cases, the "original" itself will be a reproduction. For example, it might be a reflection print that is to be reproduced again by an imaging system composed of a scanner and printer. As will be seen, each type of original has to be treated quite differently.

In discussing and working with color-imaging products and systems, it is easy to become so enamored with the technology that the real objective gets lost. It is important, then, not to forget that when it comes to images, a human observer—not a measuring instrument—is the ultimate judge of what is good or bad. Thus, regardless of the type of original being considered, one rule will remain constant throughout this book: *The assessment of color quality will be made according to the judgments of human observers.*

As obvious as that idea may seem, an experience of a colleague of ours shows that it is sometimes overlooked. He had called the manufacturer of a color-management program, purchased for his home computer, to report a problem: yellow colors always came out greenish on his monitor. The person with whom he spoke cheerfully informed him that there was no need for concern. Those greenish colors really *were* yellow; they just did not look that way because computer monitors have an overall bluish cast to them. He was told that if he were to measure those yellows, as the manufacturer had done in designing the software, he would find that they were indeed yellow. His continued protests that he "did not care how

Digital Color Management: Encoding Solutions, Second Edition Edward J. Giorgianni and Thomas E. Madden
 ISBN: 978-0-470-51244-9

they *measured*, they still *looked* greenish," were to no avail!

Since human judgments are to be the basis for determining the success or failure of color encoding and color reproduction, the basic characteristics of human color vision must be understood. These characteristics are introduced in Chapter 1, which begins with a review of color-measurement techniques that are based on the responses of a standardized representative human observer.

1

Measuring Color

Digital color encoding is, by definition, the numerical description of color in digital form. For example, in one particular color-encoding scheme, the set of digital values 40, 143, and 173 specifies a particular shade of red (the reason why will be explained later). The fact that color can be digitally encoded implies that it somehow can be measured and quantified.

But color itself is a perception, and perceptions exist only in the mind. How can one even *begin* to measure and quantify a human perception? Vision begins as light reaches the eyes; thus, a reasonable place to start is with the measurement of that light.

Light sources

In the color-science courses we often teach, students are asked to list factors they think will affect color. There usually are quite a few responses before someone mentions light sources. But perhaps this should be expected.

It is easy to take light sources, such as the sun and various types of artificial lighting, for granted. Yet unless there is a source of light, there is nothing to see. In everyday language we speak of "seeing" objects, but of course it is not the objects themselves that we see. What we see is *light* that has been reflected from or transmitted through the objects. We "prove" this in the classroom by switching off all the room lights and asking if anyone can see anything at all! This usually gets a laugh (and most often results in one or two students taking a quick nap)!

Because color begins with light, the colors that are seen are influenced by the characteristics of the light source used for illumination. For example, objects generally will look redder when viewed under a red light and greener when viewed under a green light. In order to measure color, then, it first is necessary to measure the characteristics of the light source providing the illumination.

More specifically, the *spectral power distribution* of the source, i.e., the power of its electromagnetic radiation as a function of wavelength, must be measured. Spectral power distributions can vary greatly for different types of light sources. Figure 1.1 shows, for example, the spectral power distributions for a tungsten light source and a particular type of fluorescent light source. Note that the power values in the figure are expressed in terms of relative power, not absolute power. Such relative measurements generally are sufficient for most, although not all, types of color measurements.

The most common source of light is, of course, the sun. The spectral power distribution of daylight—a mixture of sunlight and skylight—can vary greatly depending on solar altitude and on weather and atmospheric conditions. Figure 1.2 shows three of many possible examples of daylight. The undulations in each of the spectral power distributions are

Digital Color Management: Encoding Solutions, Second Edition Edward J. Giorgianni and Thomas E. Madden

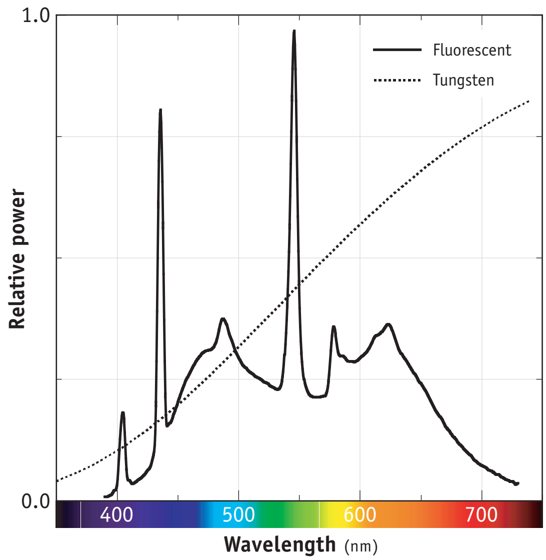


Figure 1.1 Comparison of the relative spectral power distributions for typical tungsten (dotted line) and fluorescent (solid line) light sources. The curves describe the relative power of each source's electromagnetic radiation as a function of wavelength.

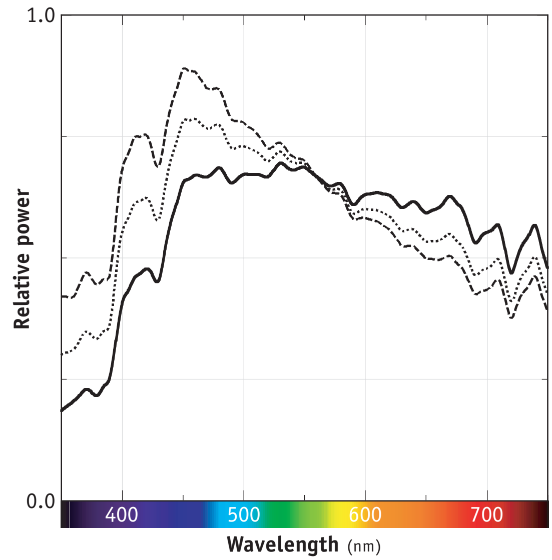


Figure 1.2 Relative spectral power distributions for three of many possible types of daylight illumination. The spectral characteristics of daylight can vary greatly depending on solar altitude, weather, and atmospheric conditions.

the result of filtration effects due to the atmospheres of the sun and the earth.

There can be a number of different light sources involved in a single digital imaging system, and each will affect the colors that ultimately are produced. For example, consider the system shown in Figure 1.3. An original scene is photographed on a color-slide film, and the slide is projected and also scanned. The scanned image is temporarily displayed on the monitor of a computer workstation, and a scan printer is used to expose a photographic paper to produce a reflection print that is then viewed.

There are six different light sources to consider in this system. First, there is the source illuminating the original scene. Another light source is used to project the slide for direct viewing. There is a light source in the slide-film scanner, which is used to illuminate the slide during scanning. The computer monitor also is a light source (the phosphors of its display emit light). The scan printer uses a light source to expose the photographic paper. Finally, a light source is used to illuminate the reflection print for viewing.

In later chapters, each of these uses of light sources will be discussed. For now, it is the measurement of color that is being discussed. Our immediate attention will be on the use of light sources to illuminate objects for viewing.

Objects

When light reaches an object, that light is absorbed, reflected, or transmitted. Depending on the chemical makeup of the object and certain other factors, the amount of light that is reflected or transmitted generally will vary at different wavelengths. For the purposes of color measurement, this variation is described in terms of spectral reflectance or spectral transmittance characteristics. These characteristics respectively describe the *fraction* of the incident power reflected or transmitted as a function of wavelength.

In most cases, an object's spectral characteristics will correlate in a straightforward way with the color normally associated with the object. For example,

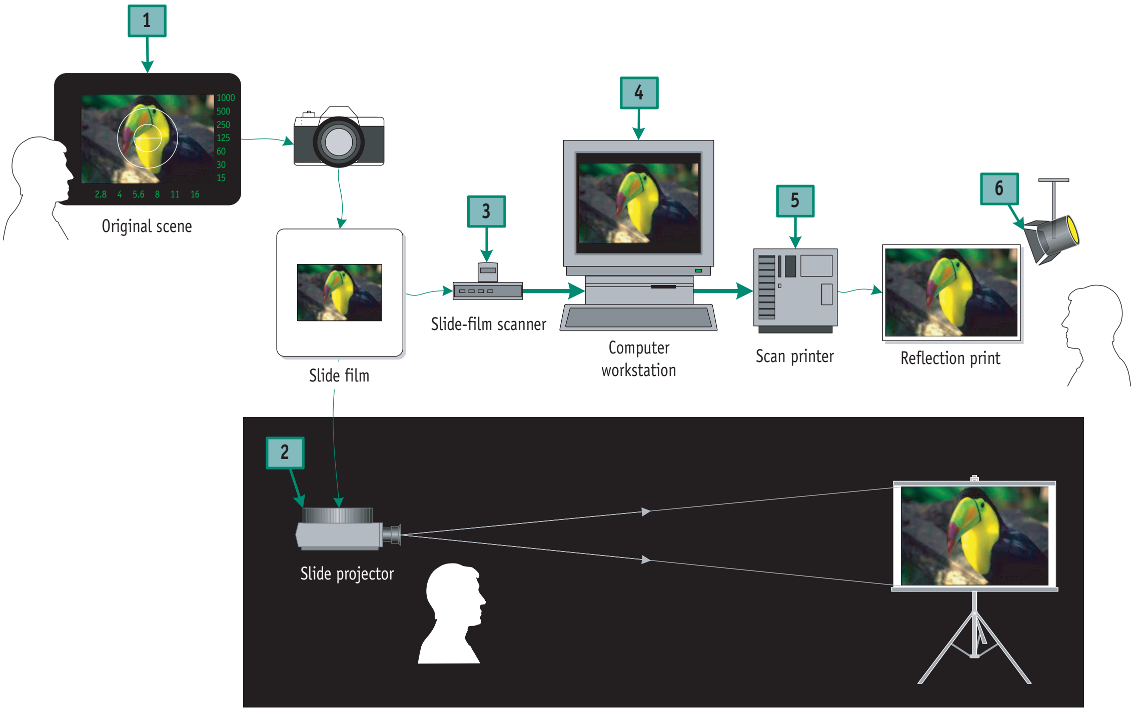


Figure 1.3 In this imaging system, there are six different light sources that contribute to the recording, reproduction, and viewing of colors.

the spectral reflectance characteristic shown in Figure 1.4 is for a red apple. The apple (generally) is seen as red because it reflects a greater fraction of red light (longer visible wavelengths) than of green light (middle visible wavelengths) or blue light (shorter visible wavelengths). Sometimes, however, the correlation of a color and its spectral reflectance characteristic is less obvious, as in the case of the two objects having the spectral reflectances shown in Figures 1.5a and 1.5b.

The object in Figure 1.5a is a particular type of flower, an ageratum. The flower appears blue to a human observer, even though it seems to have more red-light reflectance than blue-light reflectance. The object in Figure 1.5b is a sample of a dyed fabric, which appears green to a human observer, despite its unusual spectral reflectance that would seem to indicate otherwise.

In a moment, human color vision will be discussed, and the reason why these objects have color appearances that might not seem apparent from their spectral reflectances will be given. But before that can be done, it is necessary to discuss the role that objects play in the formation of what are referred to in color science as *color stimuli*.

Color stimuli

In color science, a “color” that is to be viewed or measured is referred to, more correctly, as a color stimulus. A color stimulus always consists of light. In some cases, that light might come directly from a light source itself, such as when an electronic display or the flame of a burning candle is viewed directly.

More typically, color stimuli are the result of light that has been reflected from or transmitted through various objects. For example, if the apple of Figure 1.4 is illuminated with the fluorescent light source of Figure 1.1, the resulting color stimulus will have the spectral power distribution shown in Figure 1.6.

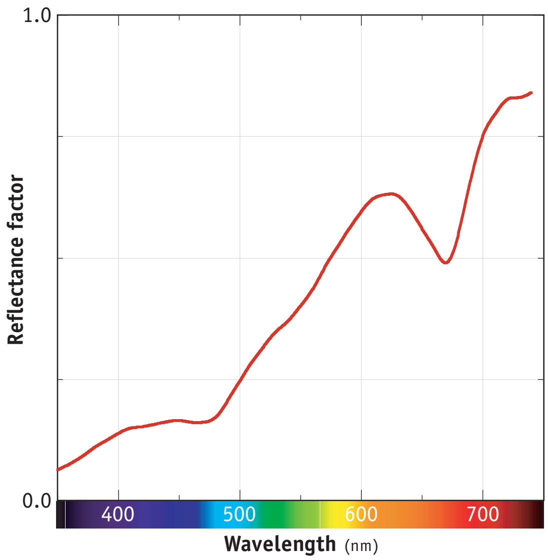


Figure 1.4 Spectral reflectance of a red Cortland apple. The apple generally is seen as red because it reflects a greater fraction of red light than of green light or blue light.

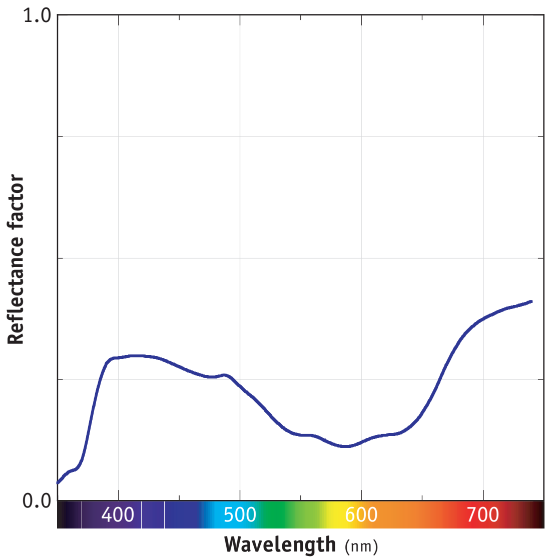


Figure 1.5a Spectral reflectance of an ageratum. The flower appears blue, even though it seems to have more red-light reflectance than blue-light reflectance.

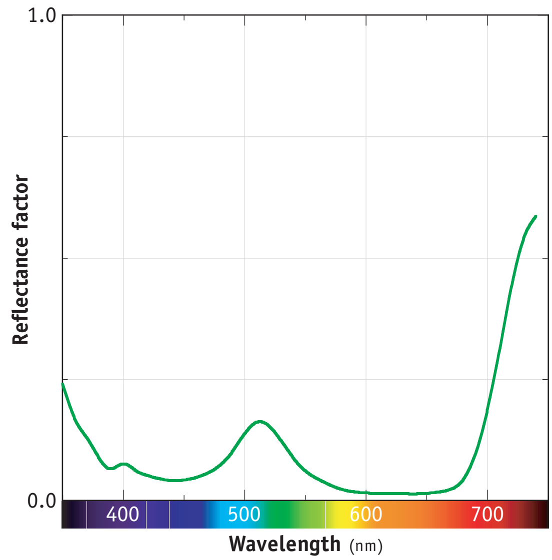


Figure 1.5b Spectral reflectance of a particular fabric sample. The fabric appears green, despite its having spectral characteristics that seem to indicate otherwise.

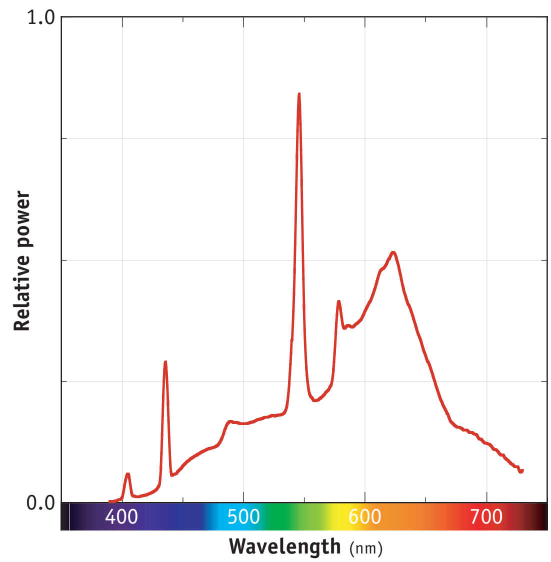


Figure 1.6 Spectral power distribution for a Cortland apple, illuminated with a fluorescent light source. In color science, such power distributions are called color stimuli.

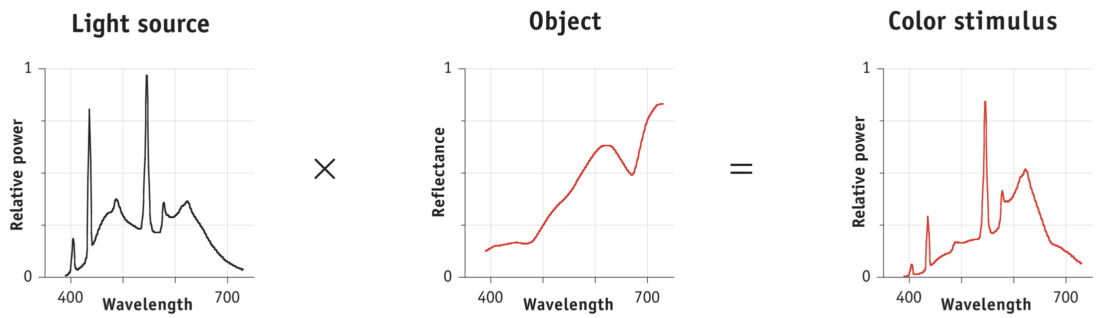


Figure 1.7 Calculation of the spectral power distribution of a color stimulus. The distribution is the product of the spectral power distribution of the light source and the spectral reflectance of the object.

The spectral power distribution of this stimulus is the *product* of the spectral power distribution of the fluorescent source and the spectral reflectance characteristic of the apple. The spectral power distribution of the stimulus is calculated simply by multiplying the power of the light source by the reflectance of the object at each wavelength, as shown in Figure 1.7.

It is important to emphasize that for a reflective or transmissive object, the color stimulus results from *both* the object and the light source. If a different light source having a different spectral power distribution illuminates an object, the color stimulus in turn will change. For example, if the apple of Figure 1.4 is illuminated with the tungsten light source of Figure 1.1, a color stimulus having the spectral power distribution shown in Figure 1.8 will be produced.

As Figure 1.8 shows, the tungsten-illuminated stimulus is very different from that produced by fluorescent illumination of the same apple. What this means is that the color of an object is *not* invariant, nor is it determined solely by the object itself. A “red” apple can be made to appear almost *any* color (or even no color at all), depending on how it is illuminated.

The concept of the color stimulus is the foundation of all methods of representing color images in numerical form. Every spatial point in a scene or image has an associated spectral power distribution. So any live scene, any image being scanned, any electronically displayed image, or any illuminated hardcopy image can be treated as a collection of individual color stimuli. These stimuli can be measured by an instrument, and they can be detected by the sensors of an imaging device.

Most importantly, it is these color stimuli that are seen by a human observer. In order to make meaningful assessments of color stimuli, then, it will

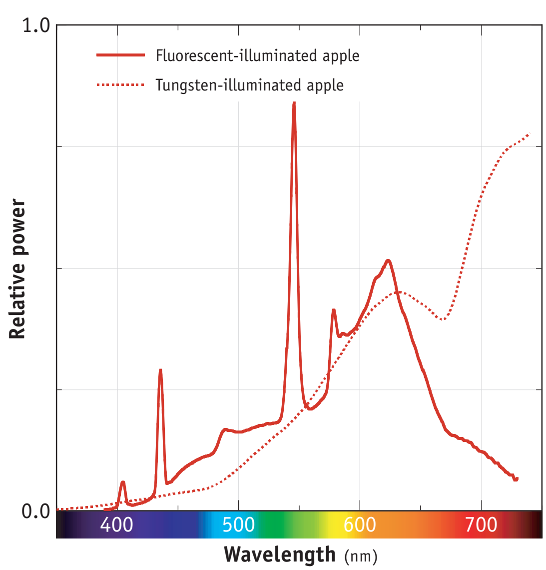


Figure 1.8 Comparison of the spectral power distributions for two stimuli—an apple illuminated by a tungsten light source (dotted line) and the same apple illuminated by a fluorescent light source (solid line).

be necessary to examine how they are detected and interpreted by the human visual system.

Human color vision

Although instruments can measure color stimuli in terms of their spectral power distributions, the eye does not interpret color stimuli by analyzing them in a comparable wavelength-by-wavelength manner. Instead, human color vision derives from the responses of just three types of photoreceptors (cones) contained in the retina of the eye. The approximate *spectral sensitivities* of these photoreceptors—their relative sensitivity to light as a function of wavelength—are shown in Figure 1.9.

Note that the sensitivity of the human visual system rapidly decreases above 650 nm (nanometers). That is why the blue flower discussed earlier appears blue, despite its reflectance at longer visible wavelengths (Figure 1.10a). The human visual system also has very little sensitivity to wavelengths below 400 nm, so the fabric discussed earlier looks green despite its high reflectances in the shorter-wavelength and longer-wavelength regions (Figure 1.10b).

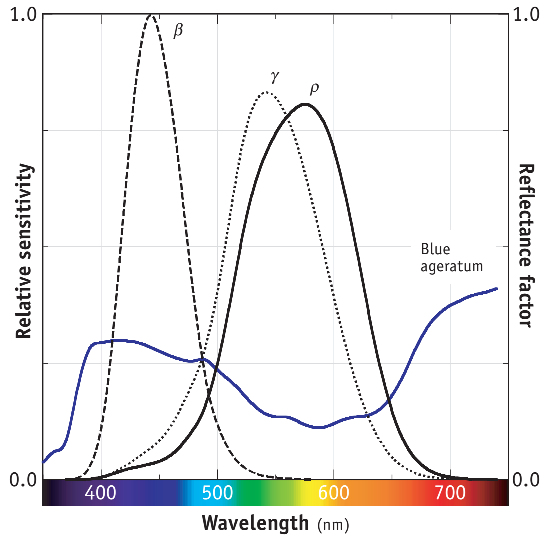


Figure 1.10a Estimated human spectral sensitivities, co-plotted with the ageratum spectral reflectance from Figure 1.5a. The sensitivity of the human visual system rapidly decreases above 650 nm, so the flower looks blue despite its reflectance at longer wavelengths.

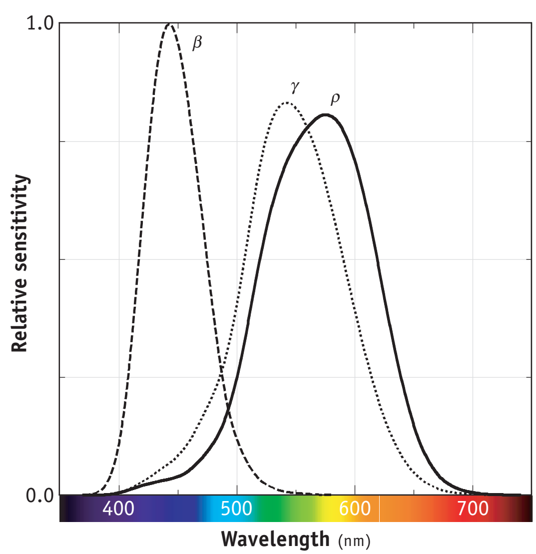


Figure 1.9 Estimated spectral sensitivities ρ, γ, and β of the three types of photoreceptors of the human eye. (The curves, derived from Estevez, 1979, have been normalized to equal area.)

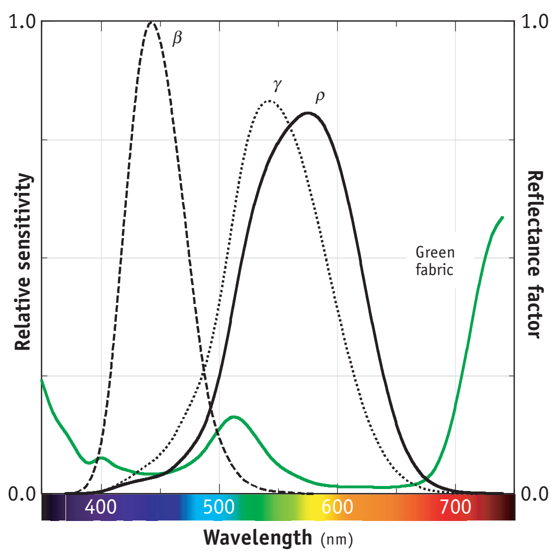


Figure 1.10b Estimated human spectral sensitivities, co-plotted with the fabric spectral reflectance from Figure 1.5b. The fabric looks green despite its high reflectances at shorter and longer wavelengths.

While this trichromatic (three-channel) analysis might seem rather inelegant, it actually is the beginning of an exquisite process that is capable of great subtlety. This process allows the human visual system to distinguish very small differences in stimulation of the three types of photoreceptors. In fact, it has been estimated that stimulation of these photoreceptors to various levels and ratios can give rise to about *10 million* distinguishable color sensations!

Because of the trichromatic nature of human vision, however, it is quite possible that two color stimuli having very different spectral power distributions will appear to have identical color. This can occur if the two color stimuli happen to produce equivalent stimulations of the photoreceptors (Figure 1.11). Two such stimuli are called a *metameric pair*, and the situation is referred to as *metamerism.*

While at first this may seem to be a problem, metamerism actually makes color-imaging systems (and digital color encoding) practical. Because of metamerism, it is not necessary either to record or to reproduce the actual spectral power distribution of

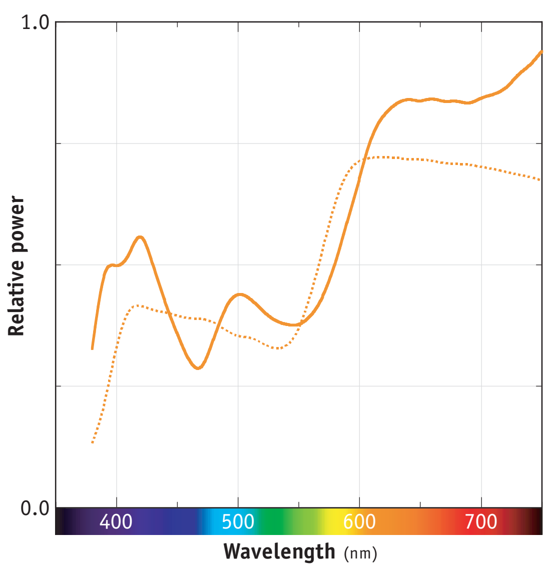


Figure 1.11 An example pair of metameric color stimuli. The two stimuli produce equivalent stimulations of the eye's photoreceptors. Metameric stimuli match in color appearance when viewed under identical conditions, but they have different spectral power distributions.

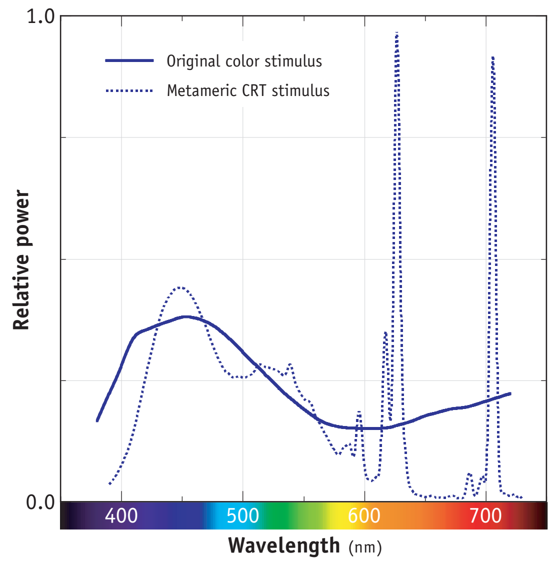


Figure 1.12 Spectral power distributions for an original color stimulus and a metameric (visually equivalent) color stimulus produced by a CRT display.

an original color stimulus. It is only necessary for an imaging system to produce a displayed stimulus that is a *visual equivalent* of the original, i.e., a stimulus that produces the same appearance. For example, in Figure 1.12 the color stimulus produced by the CRT is indistinguishable in color from the original, although its spectral power distribution obviously is very different.

As mentioned earlier, the spectral power distribution of a stimulus generally is a product of a spectral power distribution of a light source and a spectral reflectance of an object (self-luminous displays are an exception). It is important to emphasize that metamerism involves the matching of *stimuli*, not the matching of *objects*. The significance of this distinction is that two objects, having different spectral reflectances, may match metamerically under one light source, but not under another.

For example, a color copier may be capable of scanning original reflection images and producing copies that metamerically match those originals. However, if the spectral characteristics of the light source used for viewing the original images and the copies are changed the stimuli involved also will have changed, and it is likely that the copies will no longer match the originals. This is an important issue that will be revisited in later discussions on color-imaging systems and color-encoding methods.

Colorimetry

In the design of color-imaging systems and color-encoding schemes, it is important to be able to predict when two color stimuli will visually match. The science of *colorimetry* provides the basis for such predictions, and it is the foundation on which all color science is built.

Colorimetry provides methods for specifying a color stimulus by relating the measurement of its spectral power to the trichromatic responses of a defined standard observer. Doing so allows the prediction of metamerism. If two color stimuli produce the *same* trichromatic responses, those stimuli are, by definition, metameric. They will look the same (to a standard observer) if they are viewed under identical conditions.

Colorimetry is founded on a classic series of color-matching experiments that allowed the trichromatic properties of human vision to be studied and characterized. In a typical color-matching experiment, an observer views a small circular field that is split into two halves, as illustrated in Figure 1.13.

In the course of the experiment, light of a particular test color is used to illuminate one half of the field. The other half is illuminated by the superposition of light from three independent sources. *Independent* in this context means that none of the sources can be visually matched by a mixture of the other two. The independent sources (which usually are red, green, and blue) are called color *primaries*.

In performing a color-matching experiment, an observer adjusts the amounts (*intensities*) of the three color primaries until their mixture appears to match the test color. The amounts of the primaries required to produce the match are called the *tristimulus values* of the test color, for that set of color primaries. If the experiment is performed sequentially,

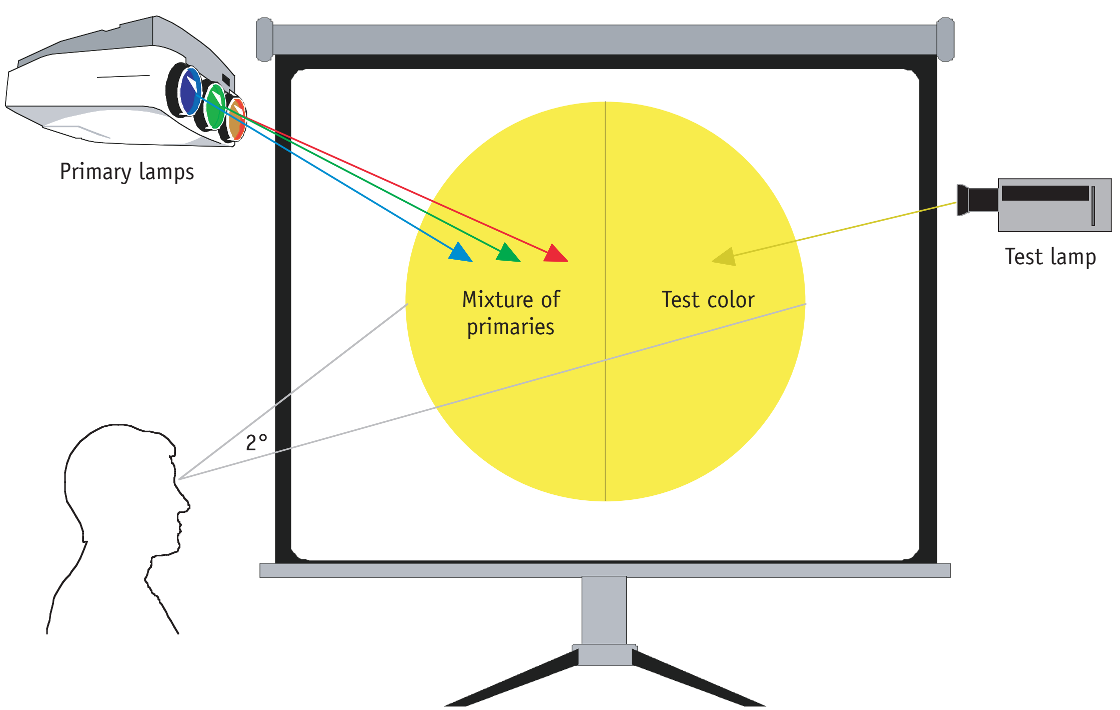


Figure 1.13 A classic color-matching experiment. A test color illuminates one half of the field, while the other half is illuminated by superposed light from three primaries. The intensities of the primaries needed to match a test color are called tristimulus values.

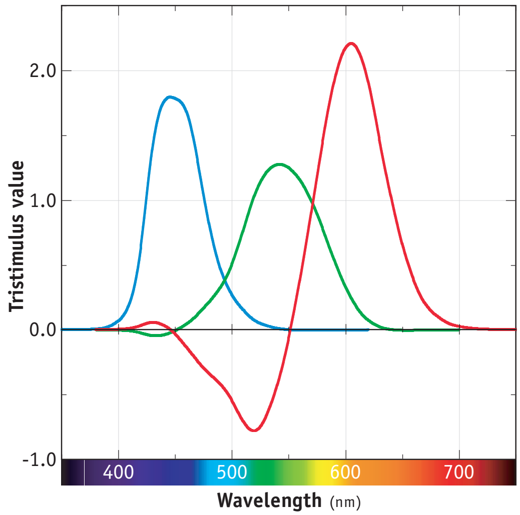


Figure 1.14 A set of color-matching functions resulting from a matching experiment performed using a particular set of red, green, and blue primaries (monochromatic light, wavelengths of 700.0, 546.1, and 435.8 nm).

using a series of test colors of monochromatic light for each of the visible wavelengths (from about 380 nm to about 780 nm), a set of three curves called *color-matching functions* is obtained. Color-matching functions represent the tristimulus values (the amounts of each of the primaries) needed to match a defined amount of light at each spectral wavelength. Figure 1.14 shows a set of color-matching functions resulting from a matching experiment performed using a particular set of red, green, and blue primaries. The color-matching functions that result will differ when different sets of color primaries are used, and they also may differ somewhat from observer to observer.

Notice that some of the tristimulus values of Figure 1.14 are *negative*. These negative values result from the fact that when the color-matching experiment is performed using monochromatic test colors, some of those test colors cannot be matched by *any* combination of the three primaries. In these cases, light from one or more of the primaries is *added* to the light of the *test color* (Figure 1.15). A match then

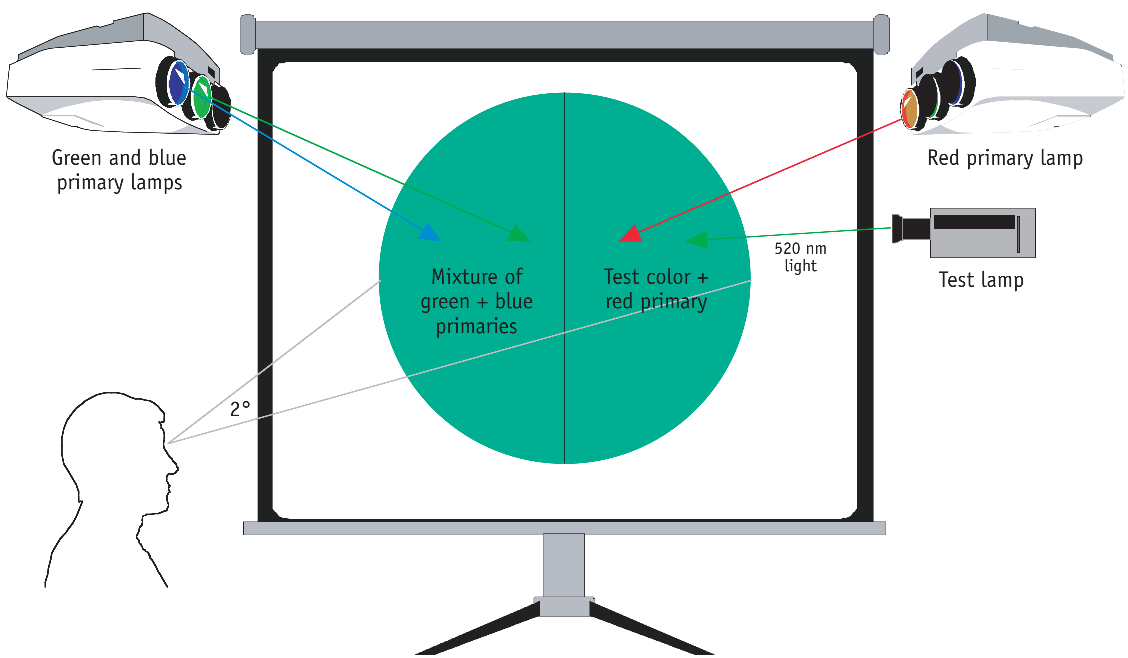


Figure 1.15 In this color-matching experiment, the test color of wavelength 520 nm cannot be matched by any combination of light from the three primaries. A match can be obtained, however, by adding light from the red primary to the test color.

can be achieved by adjusting the primaries in this configuration. Light that is added to the test color can be considered to have been *subtracted* from the mixture of the primaries. The amount of any primary added to the test color therefore is recorded as a negative tristimulus value.

It is very important to know that the color-matching functions for *any* set of physically realizable primaries will have *some* negative values. This fact will be of great significance in later discussions on a number of topics, including the signal processing requirements of color-imaging systems and the ranges of colors that can be represented by various color-encoding schemes.

The number of possible sets of color primaries is, of course, unlimited. It follows, then, that there also must be an unlimited number of corresponding sets of color-matching functions. Yet it can be shown that *all* sets of color-matching functions for a given observer are simple linear combinations of one another. A matrix operation therefore can be used to transform one set of color-matching functions to another.

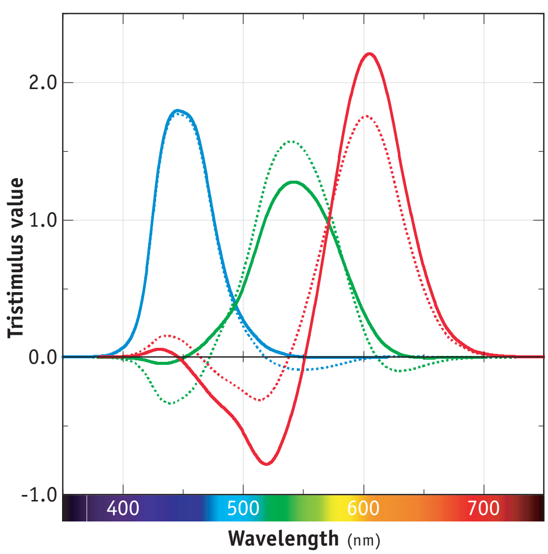


Figure 1.16 All sets of color-matching functions are linear transformations of all other sets. The set shown by the dotted-line functions $\bar{r}_2(\lambda)$, $\bar{g}_2(\lambda)$, $\bar{b}_2(\lambda)$ was derived from the set shown by the solid-line functions $\bar{r}_1(\lambda)$, $\bar{g}_1(\lambda)$, $\bar{b}_1(\lambda)$ using the linear matrix transformation given in Equation (1.1).

For the example given in Figure 1.16, the set of color-matching functions $\bar{r}_2(\lambda)$, $\bar{g}_2(\lambda)$, and $\bar{b}_2(\lambda)$ was derived from another set of color-matching functions $\bar{r}_1(\lambda)$, $\bar{g}_1(\lambda)$, and $\bar{b}_1(\lambda)$ by using the following linear matrix transformation:

$$\begin{bmatrix} \bar{r}_2(\lambda) \\ \bar{g}_2(\lambda) \\ \bar{b}_2(\lambda) \end{bmatrix} = \begin{bmatrix} 0.7600 & 0.2851 & 0.0790 \\ 0.0874 & 1.2053 & -0.1627 \\ 0.0058 & -0.0742 & 0.9841 \end{bmatrix} \begin{bmatrix} \bar{r}_1(\lambda) \\ \bar{g}_1(\lambda) \\ \bar{b}_1(\lambda) \end{bmatrix} \tag{1.1}$$

As will be seen later, this type of matrix transformation is fundamental in color science, color signal processing, and color encoding.

CIE colorimetry

In 1931 the Commission Internationale de l'Éclairage (International Commission on Illumination), the *CIE*, adopted one set of color-matching functions to define a *Standard Colorimetric Observer* (Figure 1.17) whose color-matching characteristics are representative of those of the human population having normal color vision. Although the CIE could have used any set of color-matching functions,

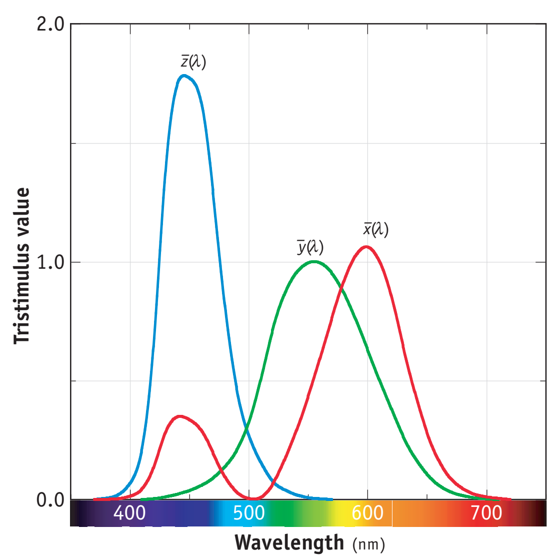


Figure 1.17 A set of color-matching functions adopted by the CIE to define a Standard Colorimetric Observer.

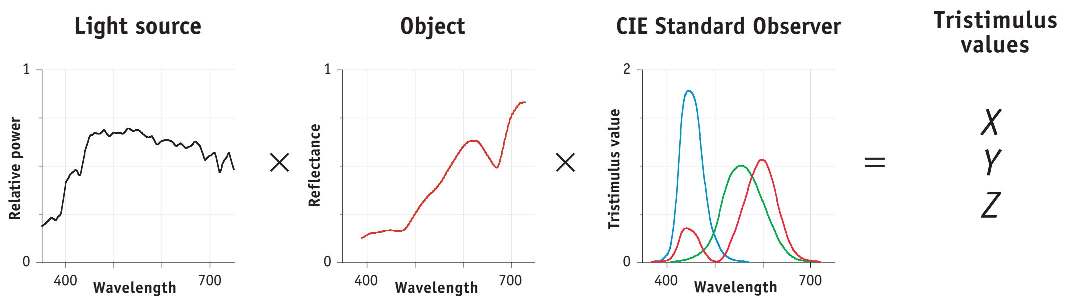


Figure 1.18 Calculation of CIE *XYZ* tristimulus values.

including a set equivalent to average ρ, γ, and β cone-response functions, this particular set was chosen for its mathematical properties.

The CIE Standard Colorimetric Observer color-matching functions are used in the calculation of *CIE tristimulus values* X, Y, and Z which quantify the trichromatic characteristics of color stimuli. The X, Y, and Z tristimulus values for a given object (characterized by its spectral reflectance or transmittance) that is illuminated by a light source (characterized by its spectral power distribution) can be calculated for the CIE Standard Colorimetric Observer (characterized by the CIE color-matching functions) by summing the products of these distributions over the wavelength (λ) range of 380 to 780 nm (usually at 5 nm intervals). This process is illustrated in Figure 1.18. The calculations of X, Y, and Z are shown in the following equations:

$$X = k \sum_{\lambda=380}^{780} S(\lambda)\, R(\lambda)\, \bar{x}(\lambda)$$
$$Y = k \sum_{\lambda=380}^{780} S(\lambda)\, R(\lambda)\, \bar{y}(\lambda) \qquad (1.2)$$
$$Z = k \sum_{\lambda=380}^{780} S(\lambda)\, R(\lambda)\, \bar{z}(\lambda)$$

where X, Y, and Z are CIE tristimulus values; $S(\lambda)$ is the spectral power distribution of a light source; $R(\lambda)$ is the spectral reflectance of a reflective object (or spectral transmittance of a transmissive object); $\bar{x}(\lambda)$, $\bar{y}(\lambda)$, and $\bar{z}(\lambda)$are the color-matching functions of the CIE Standard Colorimetric Observer; and k is a normalizing factor. By convention, k usually is determined such that $Y = 100$ when the object is a perfect white. A *perfect white* is an ideal, nonfluorescent, isotropic diffuser with a reflectance (or transmittance) equal to unity throughout the visible spectrum (Figure 1.19). *Isotropic* means that incident light is reflected (or transmitted) equally in all directions. The brightness of a perfect white therefore is independent of the direction of viewing.

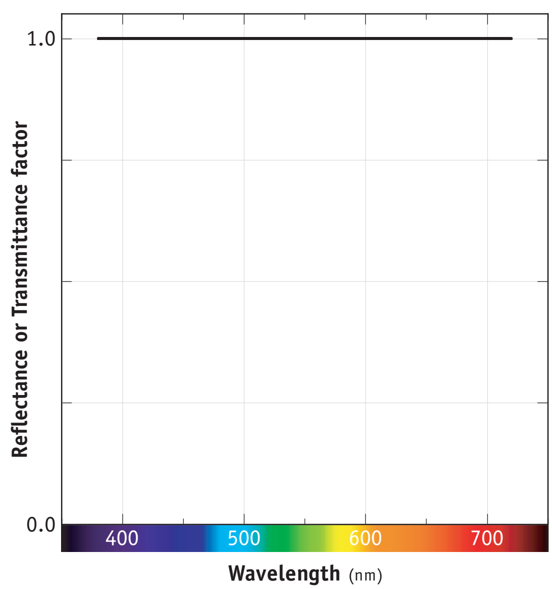


Figure 1.19 Spectral characteristic for a perfect white reflector or transmitter of light.

It was emphasized earlier that the color-matching functions for any set of physically realizable primaries will have negative values at some wavelengths. Yet the color-matching functions for the CIE Standard Colorimetric Observer (Figure 1.17) have *no* negative regions. This was accomplished by first defining a set of *imaginary* primaries and then determining the color-matching functions for those primaries. *Imaginary primaries* correspond to hypothetical illuminants having negative amounts of power at some wavelengths. For example, the imaginary "green" illuminant of Figure 1.20 has positive power in the green spectral region, but it has negative power in the blue and red regions.

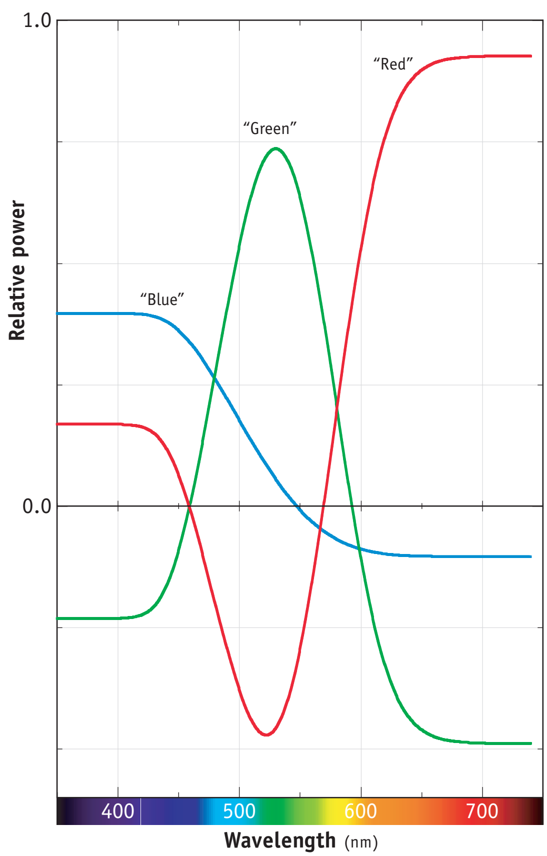


Figure 1.20 A set of spectral power distributions corresponding to a set of imaginary "red," "green," and "blue" primaries. Imaginary primaries correspond to hypothetical illuminants having negative amounts of power at some wavelengths.

While such primaries are not physically realizable, they nevertheless are very useful mathematical concepts. When they are chosen appropriately, their corresponding color-matching functions are positive at all wavelengths. Such functions are mathematically convenient because they eliminate negative values in the tristimulus calculations. (This may not seem very important today, but years ago people had to perform these calculations by hand!) Also, because the CIE Standard Colorimetric Observer color-matching functions are all positive, it is possible to construct instruments called *colorimeters*. The *spectral responsivities*—relative response to light as a function of wavelength—of a colorimeter directly correspond to the color-matching functions of the CIE Standard Colorimetric Observer. A colorimeter therefore can provide a direct measure of the CIE *XYZ* tristimulus values of a color stimulus.

Another operational convenience of the CIE color-matching functions is that the Y tristimulus value corresponds to the measurement of *luminance*. The measurement of luminance is of particular importance in color-imaging and color-encoding applications because luminance is an approximate correlate of one of the principal visual perceptions—the perception of *brightness*. When all other factors are equal, a stimulus having a higher measured luminance value will appear to be brighter than an otherwise identical stimulus having a lower measured luminance value.

Various mathematical normalizations, such as the scaling provided by the factor k in Equations (1.2), are performed in colorimetric computations. The following normalizations and definitions specifically relate to the measurement of luminance values:

- The normalization may be such that Y tristimulus values are evaluated on an absolute basis and expressed in units of luminance, typically candelas per square meter (cd/m^2). Such values are properly referred to as luminance values. However, throughout this book there will be certain instances when it is particularly important to emphasize that absolute, not relative, amounts of light are being referred to. In these instances, the somewhat redundant expression *absolute luminance values* will be used to provide that emphasis.

- When the normalization is such that the Y tristimulus value for a perfect white object is 1.0, normalized Y values are called luminance-factor values.
- When the normalization is such that the Y tristimulus value for a perfect white object is 100, normalized Y values are called *percent luminance-factor values.*

Although the X and Z tristimulus values have no direct perceptual correlates, they are used in the calculation of tristimulus-value ratios called *chromaticity coordinates*. The chromaticity coordinates x, y, and z describe the qualities of a color stimulus apart from its luminance. They are derived from the tristimulus values as follows:

$$x = \frac{X}{X+Y+Z}$$
$$y = \frac{Y}{X+Y+Z} \qquad (1.3)$$
$$z = \frac{Z}{X+Y+Z}$$

A plot of y versus x is called a *chromaticity diagram* (Figure 1.21a). The horseshoe-shaped outline is the *spectrum locus*, which is a line connecting the points representing the chromaticities of the spectrum colors. In the figure, the ends of the spectrum locus are connected by a straight line known as the *purple boundary*. The chromaticity coordinates of *all* physically realizable color stimuli lie within the area defined by the spectrum locus and the purple boundary.

Figure 1.21b shows the locations of the chromaticity coordinates of the real *RGB* primaries corresponding to the color-matching functions of Figure 1.14. The triangle formed by connecting those locations encloses the chromaticity coordinates of all color stimuli that can be matched using positive (including zero) amounts of those real primaries. Also indicated are the locations of the chromaticities of the imaginary *XYZ* primaries corresponding to the color-matching functions of the CIE Standard Colorimetric Observer. Note that the triangle formed by connecting those locations encloses the entire area defined by the spectrum locus and the purple boundary. That is why all real color stimuli can be matched using positive amounts of those imaginary primaries.

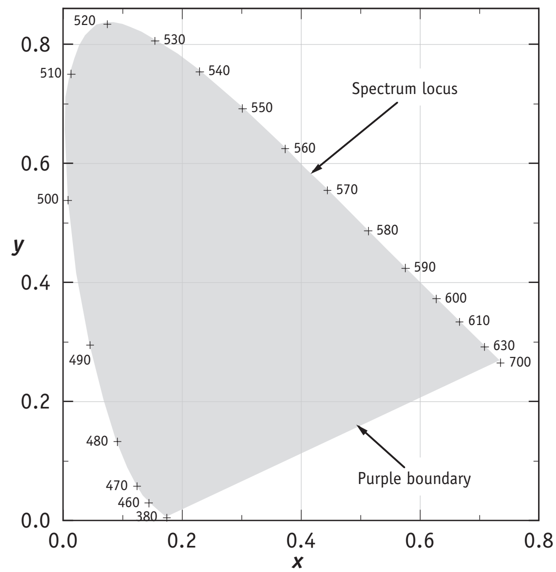


Figure 1.21a CIE x, y chromaticity diagram. The chromaticity coordinates define the qualities of a color stimulus apart from its luminance. The chromaticity coordinates of all physically realizable color stimuli lie within the area defined by the spectrum locus and the purple boundary.

The CIE also has recommended the use of other color-coordinate systems, derived from *XYZ* in which perceived differences among colors are represented more uniformly than they are on an x, y chromaticity diagram. These recommendations include the CIE 1976 u', v' Metric Chromaticity Coordinates and the CIE 1976 $L^*a^*b^*$ (CIELAB) and CIE 1976 $L^*u^*v^*$ (CIELUV) color spaces. (Refer to Appendix A for more details regarding these color spaces.)

All of the CIE coordinate systems are quite useful for specifying small color *differences* between color stimuli (CIE colorimetry was, in fact, developed specifically for that purpose), but it is essential to understand that *none of these systems specifies the appearance of those stimuli.* The reason for this will be discussed in Chapter 3. For now, the reader should

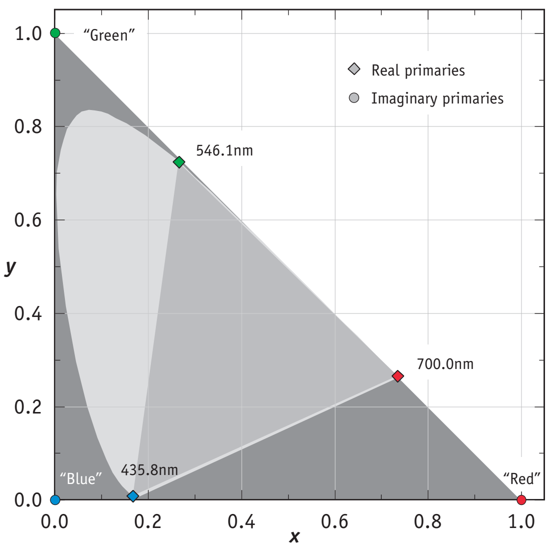


Figure 1.21b Chromaticity coordinates of the real primaries corresponding to the color-matching functions of Figure 1.14, and those of the imaginary primaries corresponding to the color-matching functions of Figure 1.17. The triangle formed by connecting the chromaticity coordinates of a set of primaries encloses the chromaticity coordinates of all color stimuli that can be matched using all-positive amounts of those primaries.

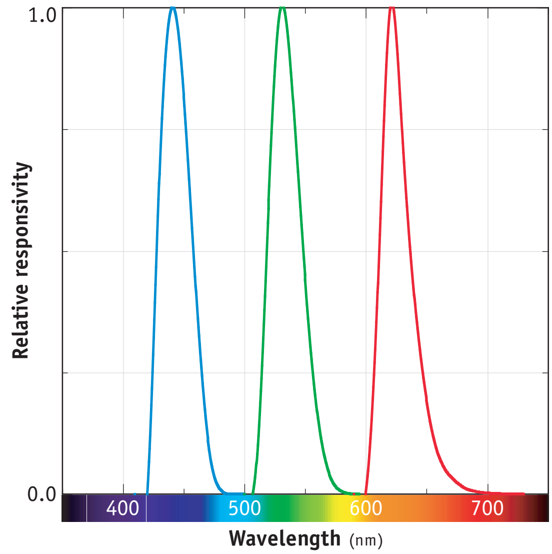


Figure 1.22a Red, green, and blue spectral responsivities for an ISO Status A densitometer.

be cautioned that this fact is commonly misunderstood and, as a consequence, CIE coordinate systems frequently are misinterpreted and misused as if they describe color appearance. The distinction between colorimetry and color appearance may seem subtle and of interest only to color scientists. In practice, however, failures to recognize that distinction have been responsible for the demise of numerous color-encoding methods, color-management systems, and entire color-imaging systems.

Other color measurements

In addition to CIE colorimetry, there are other types of color measurements that are relevant to color imaging and color encoding. Of particular importance in imaging applications involving hardcopy media is the measurement of *optical density*. Optical density values can be measured using instruments called *densitometers*. (Refer to Appendix B for more details regarding densitometers and densitometric measurements.)

Optical densities of color media generally are measured using three-channel (or sometimes four-channel) densitometers. The spectral responsivities for these instruments are defined by industry standards. Figure 1.22a, for example, shows the specified red, green, and blue spectral responsivities for an ISO Standard Status A densitometer.

Status A densitometers are widely used for measurements of photographic media and other types of hardcopy media that are meant to be viewed directly by an observer. That fact might seem to suggest that Status A density values must provide information equivalent to that provided by CIE colorimetric values, but that is not the case. Status A spectral responsivities do not correspond to those of a CIE Standard Observer, as shown in Figure 1.22b, or to any other set of visual color-matching functions. As a result, two spectrally different objects that metamerically match under a particular illuminant (i.e., their stimuli have the same CIE *XYZ* values) are unlikely to have matched red, green, and blue Status A values.

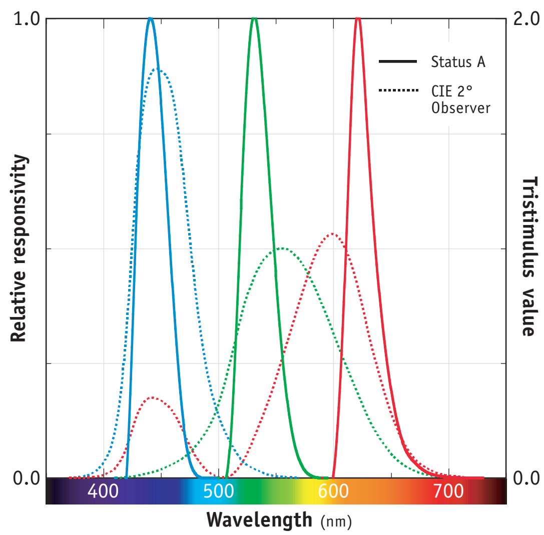


Figure 1.22b Comparison of ISO Status A spectral responsivities to the color-matching functions of the CIE Standard Colorimetric Observer (normalized to equal area).

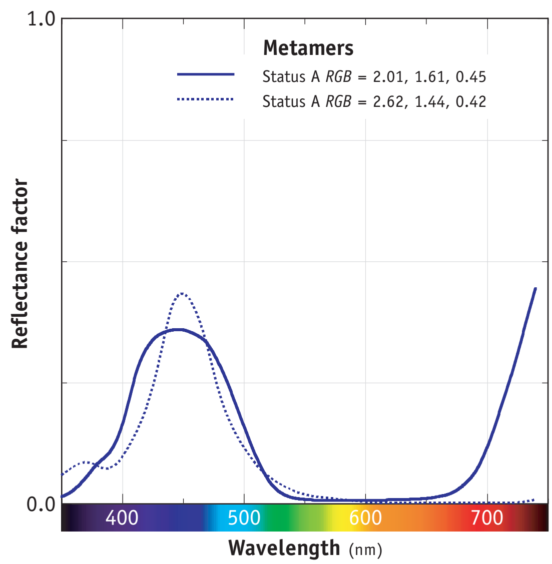


Figure 1.23 Spectral reflectances of two objects that metamerically match under a particular illuminant but that have very different *RGB* Status A density values.

For example, Figure 1.23 shows the spectral reflectances of two objects. Although these objects metamerically match under a particular illuminant, their red, green, and blue Status A density values *differ* by −0.61, 0.17, and 0.03 respectively. The converse also is true: a pair of objects having the same red, green, and blue Status A values is unlikely to be a metameric pair. Figure 1.24 shows the spectral reflectances of two different objects. While these objects have identical Status A densities, their X, Y, and Z tristimulus values, under a particular illuminant, differ by 15.7, 14.6, and −0.8 respectively.

Film scanners and reflection scanners (Figure 1.25) are used for color measurement of hardcopy input images on digital color-imaging systems. While various sets of red, green, and blue spectral responsivities are used in different types of color scanners, those responsivities seldom correspond to a set of color-matching functions. Most scanners, therefore, essentially are *densitometers*; they are *not* colorimeters. As will be shown later, that is a critical distinction that must be taken into account in the digital encoding of scanned colors.

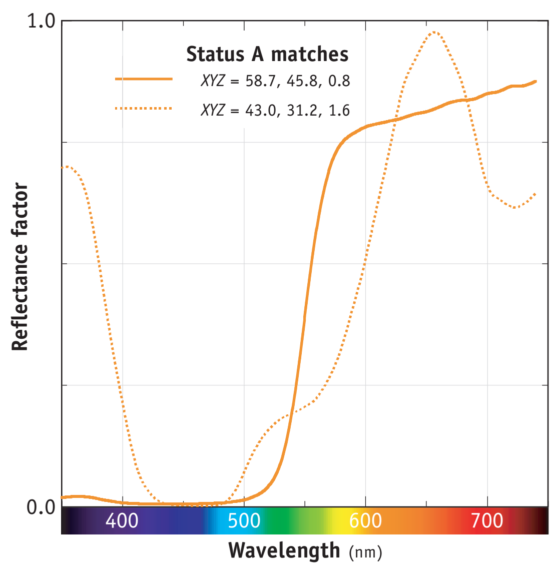


Figure 1.24 Spectral reflectances of two objects that have identical *RGB* Status A density values but that do not match metamerically.

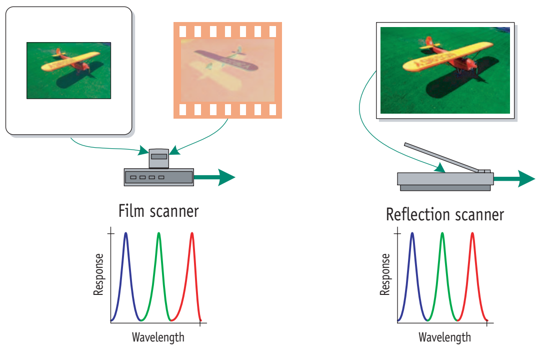


Figure 1.25 A film scanner and a reflection scanner. The spectral responsivities of most scanners do not correspond to a set of color-matching functions. Therefore, most scanners are densitometers, not colorimeters.

Stimuli produced by self-luminous display devices, such as CRTs, often are measured in terms of light intensity. Three-channel instruments, which are somewhat similar to densitometers, can make simultaneous readings of red-light, green-light, and blue-light intensities. Various types of single-channel instruments also can be used in the measurement of self-luminous displays. When such instruments are used, separate red-light, green-light, and blue-light readings can be made by sequentially sending individual red, green, and blue signals to the display device that is to be measured.

Summary of key issues

- All vision is a response to light.
- Light sources are characterized by their spectral power distributions.
- Objects are characterized by their spectral reflectances or transmittances.
- Color stimuli generally are produced by a light source and an object; stimuli are characterized by their spectral power distributions.
- Scenes and images are collections of individual color stimuli.
- Human color vision is trichromatic.
- Metameric color stimuli differ spectrally but match in appearance.
- CIE colorimetry allows the prediction of metameric matching between color stimuli; metameric colors have the same CIE *XYZ* tristimulus values.
- CIE colorimetry was developed for specifying the trichromatic properties of color stimuli.
- CIE colorimetric values can indicate how much the appearance of two stimuli will differ, if the differences in their trichromatic properties are sufficiently small.
- CIE colorimetric values such as CIE *XYZ* tristimulus values, CIE $L^*a^*b^*$ values, and CIE $L^*u^*v^*$ values do not describe color appearance.
- Densitometers are used to measure the optical densities of hardcopy media. Their spectral responsivities do not correspond to any set of visual color-matching functions.
- A colorimeter directly measures CIE *XYZ* tristimulus values.
- Most image scanners essentially are densitometers and not colorimeters; their spectral responsivities generally do not correspond to any set of visual color-matching functions.

2

Color-Imaging Systems

In Chapter 1 it was shown that, for the purposes of color measurement, scenes and other images can be characterized in terms of their color stimuli. In this chapter, the fundamental principles of how such color stimuli can be captured and reproduced by color-imaging systems will be discussed.

Color-imaging systems can be built using an almost unlimited variety of optical, chemical, and electronic components. But regardless of what technologies they incorporate, all imaging systems must perform three basic functions: *image capture*, *signal processing*, and *image formation* (Figure 2.1). These functions are the building blocks of all color-imaging systems, from the simplest to the most complex.

Image capture

To form a reproduction, an imaging system first must detect light from each original color stimulus and, from that light, produce a detectable image signal. This function, called *image capture*, can be realized in a number of ways, depending on the technology of the particular imaging system.

For example, an electronic camera, such as a digital still camera, might use a solid-state image sensor, such as a *charge-coupled device* (CCD), to detect light. Image capture occurs as photons of light are absorbed by the sensor, resulting in the generation of electrons. These electrons are collected into charge packets, and an image signal is produced by a sequential readout of those packets.

In a photographic film, light is captured in the form of a *latent image*. The latent image, composed of small clusters of metallic silver, is produced as photons of light strike the light-sensitive silver halide grains of the film. This chemical signal is detected and amplified during subsequent chemical processing. (Refer to Appendix C for more details on photographic media.)

Accurate color reproduction requires image capture that is, like the human eye, trichromatic. As part of the capture process, then, the spectral content of original color stimuli must be separated to form three distinguishable color signals. This generally is accomplished by some form of trichromatic image capture. In some special applications, however, more than three color channels are captured and trichromatic color signals are subsequently derived.

In an electronic camera, trichromatic capture can be accomplished using a solid-state sensor composed of a mosaic of light-sensitive elements. Individual sensor elements are overlaid with red, green, or blue filters (Figure 2.2a). Some high-end video and digital motion picture cameras use three sensors (for higher spatial resolution) and an appropriate arrangement of beam splitters and color filters (Figure 2.2b).

Trichromatic capture is accomplished in color photographic media by the use of overlaid light-sensitive layers. In the simplified film cross-section

Digital Color Management: Encoding Solutions, Second Edition Edward J. Giorgianni and Thomas E. Madden
 ISBN: 978-0-470-51244-9

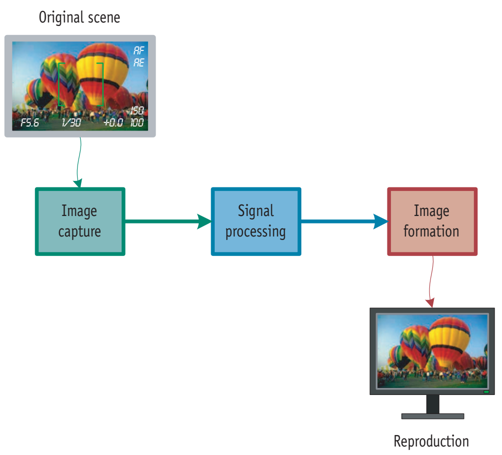


Figure 2.1 These three basic functions are the building blocks of all color-imaging systems, from the simplest to the most complex.

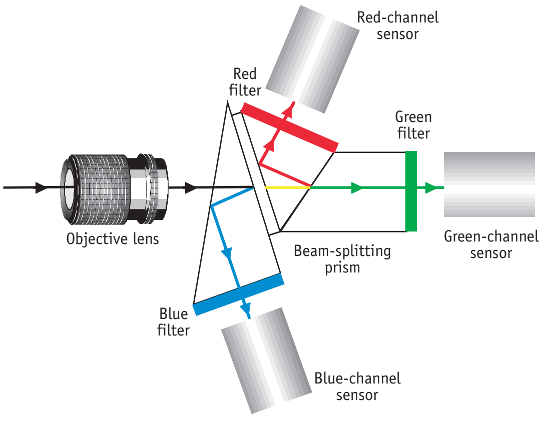


Figure 2.2b Use of three color-separation filters for trichromatic image capture in a three-sensor video camera.

shown in Figure 2.3, the top layer records blue light, the middle layer records green light, and the bottom layer records red light. A color film actually may contain a total of 12 or more image-forming and other special purpose layers, but its behavior is fundamentally the same as that of a simple three-layer film.

The color characteristics of trichromatic capture are determined by the *spectral responsivities* (relative responses to light as a function of wavelength) of the particular system. These responsivities typically will differ among systems. For example, Figure 2.4

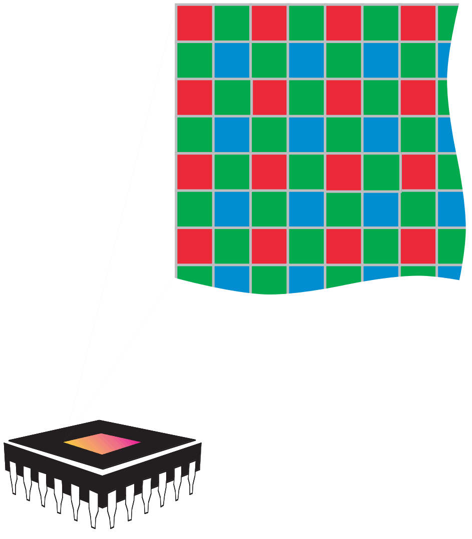


Figure 2.2a Trichromatic image capture can be achieved using a CCD sensor with an integral mosaic of red, green, and blue filters. In this example, the filters are arranged in a Bayer pattern.

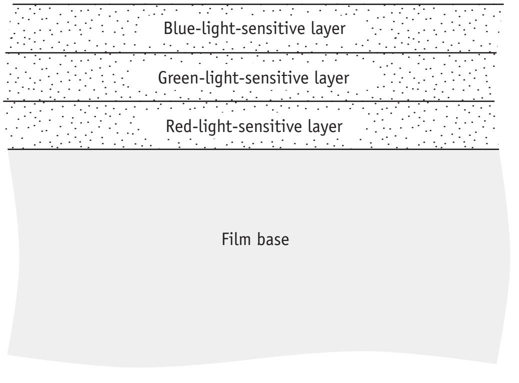


Figure 2.3 A simplified cross-section of a photographic film. Trichromatic image capture is achieved by the use of three layers sensitive to red, green, or blue light.

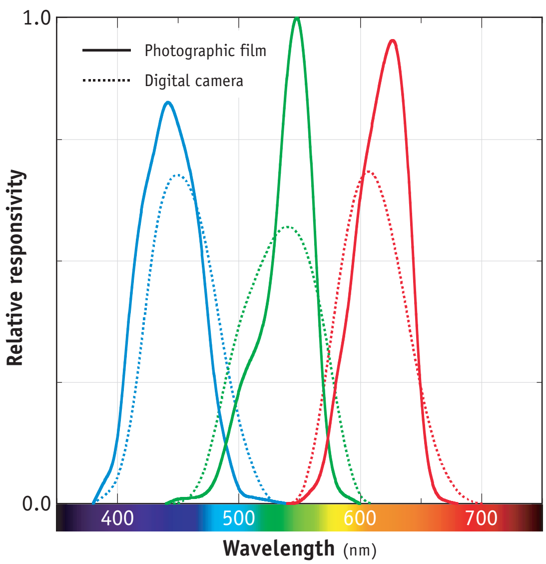


Figure 2.4 Comparison of spectral responsivities of a particular digital camera and a particular photographic film.

compares the red, green, and blue spectral responsivities for a particular photographic film and a particular digital camera.

The individual color responses, called *exposures*, produced by a given set of spectral responsivities can be calculated using the following equations:

$$
\begin{aligned}
R_{exp} &= k_{c_r} \sum_{\lambda} S(\lambda)\, R(\lambda)\, r_c(\lambda) \\
G_{exp} &= k_{c_g} \sum_{\lambda} S(\lambda)\, R(\lambda)\, g_c(\lambda) \\
B_{exp} &= k_{c_b} \sum_{\lambda} S(\lambda)\, R(\lambda)\, b_c(\lambda)
\end{aligned}
\tag{2.1}
$$

where R_{exp}, G_{exp}, and B_{exp} are red, green, and blue exposure values; $S(\lambda)$ is the spectral power distribution of the light source; $R(\lambda)$ is the spectral reflectance (or transmittance) of the object; $r_c(\lambda)$, $g_c(\lambda)$, and $b_c(\lambda)$ are the red, green, and blue spectral responsivities of the image-capture device or medium; and k_{c_r}, k_{c_g}, and k_{c_b} are normalizing factors. These factors usually are determined such that R_{exp}, G_{exp}, and $B_{exp} = 1.0$ when the object is a perfect white.

This white-point normalization is the equivalent of performing a *white balance* adjustment on an electronic camera, where the red, green, and blue image-capture signals are adjusted independently such that equal *RGB* reference voltages are produced when a reference white object is imaged or measured. With the application of the three normalization factors, the computed exposure values become *relative* values, properly referred to as *exposure-factor values.*

Equations (2.1) essentially are the same in form as those used for computing CIE *XYZ* tristimulus values (Chapter 1, Equations (1.2)). In fact, if the red, green, and blue spectral responsivities of the image-capture stage of an imaging system corresponded to the color-matching functions of the CIE Standard Colorimetric Observer, the resulting *RGB* exposure-factor values would be equivalent to CIE *XYZ* tristimulus values. In other words, the image-capture device or medium essentially would be a colorimeter.

This raises an interesting question. For accurate color reproduction, should the spectral responsivities of an imaging system *always* be designed to match those of a standard human observer? The answer is not as straightforward as it might seem. This question will be revisited in Part II, where a closer look will be taken at the second basic function performed by all imaging systems: *signal processing*.

Signal processing

Signal processing modifies image signals produced from image capture to make them suitable for producing a viewable image. For example, an image-capture signal produced by a broadcast video camera is electronically processed and amplified for transmission. A home television receiver performs further signal processing to produce signals appropriate for driving its particular type of display. In a photographic film, signal processing occurs as the film is chemically processed. (Chemical photographic processing sometimes is referred to as "developing." However, image development actually is just one of several steps in the overall chemical process, so the term is not strictly correct and will not be used here.)

Signal processing typically includes linear and nonlinear transformations operating on the individual color signals. A linear transformation of an individual color signal might effect a simple amplification, which usually is required to produce a signal sufficiently strong to generate a viewable image. Nonlinear transformations of the individual color signals primarily are used to control the *grayscale* of the image produced by the system. The grayscale is a measure of how a system reproduces a series of *neutral* colors, ranging from black to white. In Part II, it will be shown that the grayscale characteristic is one of the most important properties of any imaging system, and the reasons why the signal processing associated with the grayscale must be highly nonlinear will be discussed.

Signal processing also is used to create linear and nonlinear interactions (*crosstalk*) among the individual color signals. For example, a modified red signal might be formed by taking portions of the green and blue signals and adding them to or subtracting them from the original red signal. In Part II, the reasons why such interactions are necessary will be discussed. Signal processing also may include spatial operations such as image compression, sharpening, and noise reduction. While spatial operations may not directly affect color, they are a factor that must be considered in developing the most appropriate form of digital color encoding for a given system.

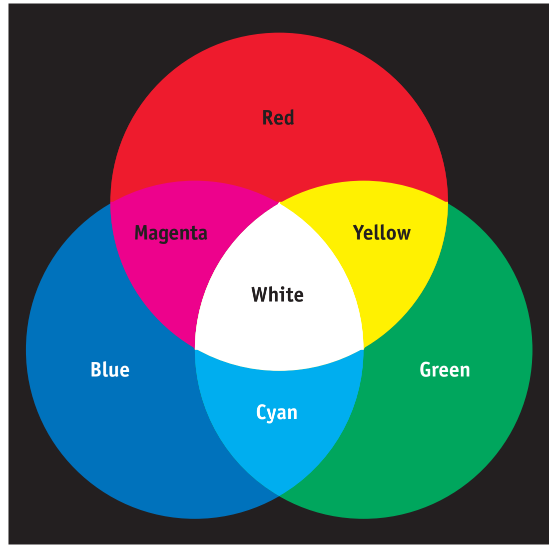


Figure 2.5 Additive mixing of red, green, and blue additive color primaries. Mixing red and green forms yellow; mixing red and blue forms magenta; and mixing green and blue forms cyan. Mixing full intensities of red, green, and blue forms white.

Image formation

The ultimate goal of a color-imaging system is, of course, to produce a viewable image. That is accomplished by the final stage of the system, *image formation*, where processed image signals are used to control the color-forming elements of the output medium or device. Although there are many types of color-imaging media, color-imaging devices, and color-image-forming technologies, virtually all practical image-formation methods fall into one of two basic categories: additive color or subtractive color.

In the formation of most *additive color* images, processed image signals are used to directly control the intensities of primary colored lights that make up the displayed image. Colors are produced by additive mixing (Figure 2.5). A three-beam digital cinema projector, for example, forms images by additively combining modulated intensities of red, green, and blue lights on a projection screen. This directly generates color stimuli, so no other light source is required for viewing. Color CRTs, LCD panels, plasma panels, and similar direct-view display devices form images by generating adjacent pixels of red, green, and blue light. The mixing of that red, green, and blue light subsequently takes place within the visual system of the observer.

In the formation of *subtractive color* images, processed image signals control the amounts of three or more *colorants* (dyes, inks, or pigments) that selectively *absorb* (subtract) light of specific wavelength regions (Figure 2.6). Photographic media, for example, use *cyan*, *magenta*, and *yellow (CMY)* image-forming dyes to absorb red, green, and blue light, respectively. Many printing processes use CMY inks plus an additional black ink (the *K* of *CMYK*). An image formed by subtractive colorants is an *object*, which requires a light source for viewing. The spectral properties of the color stimuli produced from a subtractive image will change if the spectral power distribution of the viewing light source is changed.

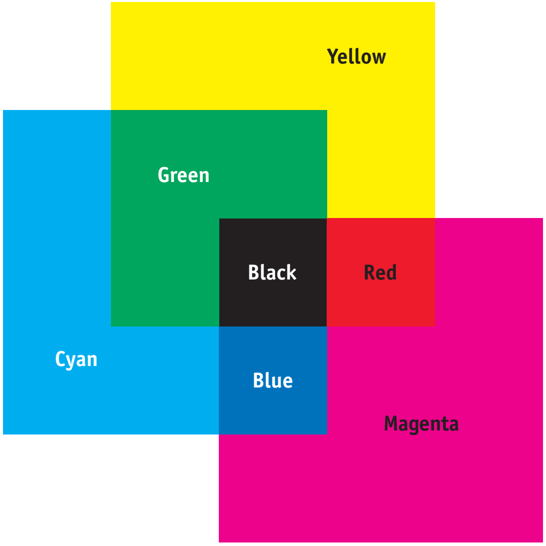


Figure 2.6 Subtractive mixing of cyan, magenta, and yellow color primaries. Mixing cyan and magenta forms blue; mixing cyan and yellow forms green; and mixing magenta and yellow forms red. Mixing maximum amounts of cyan, magenta, and yellow forms black.

Complete color-imaging systems

A complete imaging system can be defined as any combination of devices and/or media capable of performing all three basic functions described in this chapter. For example, the combination of a digital still camera (image capture), a computer (signal processing), and a monitor (image formation) forms a complete system. Somewhat less obvious is that, by this definition, photographic media—such as color-slide and negative films—also are complete imaging systems. That fact will be extremely important to remember later when the problems of color encoding in hybrid systems, which combine photographic media, other types of hardcopy media, and various types of electronic devices, are addressed.

Summary of key issues

- All imaging systems must perform three basic functions: image capture, signal processing, and image formation.
- An imaging system can be any combination of devices and/or media that is capable of performing these three functions.
- Color imaging requires image capture that is (at least) trichromatic.
- Signal processing modifies image signals, produced by image capture, to make them suitable for forming a viewable image.
- In the image-formation stage, processed image signals control the amounts of the color-forming elements produced by the output device or medium.
- Color-image formation can be accomplished either by additive or by subtractive color techniques.

3

The Human Color-Imaging System

In Chapter 2, color-imaging systems were described in terms of the three basic functions they must perform. In this chapter, those three functions will be used to examine the most important color-imaging system of all: the visual system of the human observer. In this discussion, and throughout the book, the visual system will be dealt with in the context shown in Figure 3.1.

The figure is meant to illustrate that both the original and its reproduction are to be viewed, and that a human observer will judge the color quality of the reproduction. This is a common situation, and it conforms to the basic rule established in the introduction to this section, but it does not seem like a very scientific method of assessing color quality. Would it not be better to replace the observer with some type of objective measurement instrument? And, if so, what kind of instrument should be used?

Colorimeters directly measure CIE *XYZ* tristimulus values of color stimuli, so a colorimeter would certainly seem an obvious choice as a substitute for the observer. It could be used to measure the *XYZ* values of both the original and the reproduction (Figure 3.2). The quality of the reproduction then could be quantified in terms of how well its *XYZ* values (or other colorimetric values, such as CIELAB $L^* a^* b^*$ values, derived from the *XYZ* values) match those of the original.

As logical as this approach might seem, it is fundamentally flawed. To understand why, one has to consider what goes on when an observer judges the quality of a reproduction. In most cases, color quality will be judged based on a comparison of the image that forms in the mind of the observer, when viewing the reproduction, to some reference mental image. This reference image may derive from memory, if the observer actually saw the original at some time, or from the observer's conception of how the original should appear, based on his or her personal experience. In other cases, the observer may be looking at both the original and the reproduction at the same time, so the reference and reproduction mental images will form simultaneously.

In any of these situations, the judgment will be based on a comparison of mental images that are the *end product* of the visual system. To produce these images, the human visual system must perform the same basic functions required of any other color-imaging system.

Figure 3.3 shows the human visual system in terms of those functions. The *image-capture* process is performed by the trichromatic responses of the eye. The *image-formation* process corresponds to the formation of a perception of a color image in the observer's mind. The formation of this mental image is influenced by intermediary visual *signal processing*, which can be broadly divided into two types: psychological and psychophysical.

Psychological signal processing includes effects due to color memory, which generally is not

Digital Color Management: Encoding Solutions, Second Edition Edward J. Giorgianni and Thomas E. Madden
 ISBN: 978-0-470-51244-9

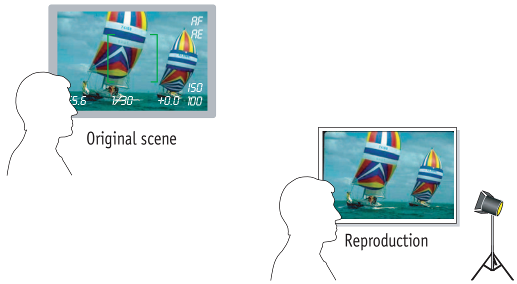


Figure 3.1 An original scene and its reproduction, where both are viewed and evaluated by a human observer.

colorimetrically accurate. It also includes color preference, which can be different from color memory, and various cognitive effects that cause the observer to perceive colors somewhat according to expectations and experience.

Psychophysical signal processing includes a variety of effects that are due to both physiological and mental processes. These effects result from various forms of visual adaptation, which are discussed next.

Adaptation

Adaptation refers to the process by which the visual mechanism adjusts to the conditions under which the eyes are exposed to radiant energy. The relationship between the physical characteristics of a color stimulus and the perception of its color is strongly

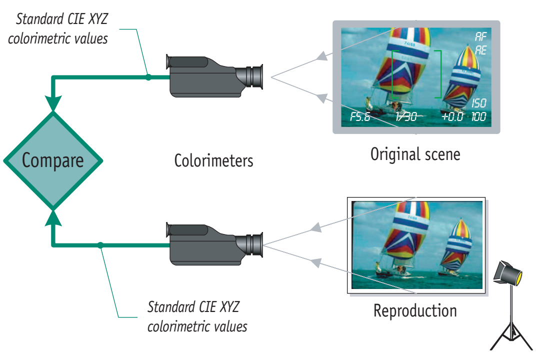


Figure 3.2 An original scene and its reproduction. Both are measured by a colorimeter, and the resulting colorimetric values are compared.

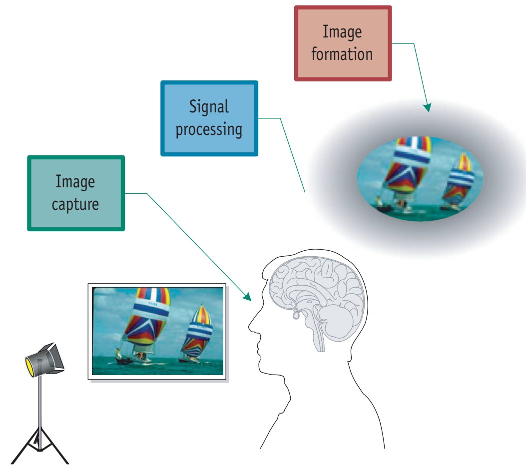


Figure 3.3 The human visual system can be described in terms of the same basic functions used to describe all other color-imaging systems.

influenced by effects produced by various forms of adaptation.

In the design of color-imaging media and systems, there are three adaptation effects of particular importance: general-brightness adaptation, lateral-brightness adaptation, and chromatic adaptation.

General-brightness adaptation refers to the adjustments of the visual mechanism in response to the overall level of the stimulus (or collection of stimuli) to which the eyes are exposed. For example, when the eyes are exposed for a sufficient length of time to a low level of illumination, the visual receptors compensate by becoming relatively more sensitive.

The human visual system, then, works on a *relative* basis, not on an absolute basis. This allows the recognition of objects under a wide range of illumination conditions. For example, a white flower likely will be seen as white whether it is viewed indoors or outdoors, even though the *absolute* amount of light reflecting from the flower is very different in the two viewing environments.

Despite this adaptation, visual perceptions still are affected somewhat by the absolute luminance levels of the stimuli being viewed. Stimuli having lower absolute luminance levels first of all are perceived as being somewhat lower in *luminance contrast*. This means that the light and dark *differences* among stimuli of higher and lower luminances are less apparent. In addition, stimuli having lower absolute luminance levels have less *colorfulness*. This means that color stimuli appear to exhibit less of their particular hue.

For example, consider a reflection print viewed indoors, where the image luminances are low because of a low level of illumination, and also viewed outdoors, where a higher level of illumination results in higher image luminances. The print will appear quite different in the two environments, even to an observer who has become adapted to each of those environments. When the print is viewed indoors, areas of black will look less black, areas of white will look less white, and colors generally will look less colorful.

Lateral adaptation refers to changes induced in the sensitivities of adjacent areas of the retina. For example, the sensitivity of a particular retinal receptor may be increased or decreased depending on the amount of light being received by neighboring receptors. This important perceptual effect is called *lateral-brightness adaptation*. Lateral effects help the visual system to discriminate objects by making their edges more apparent.

Another manifestation of lateral-brightness adaptation is that the apparent luminance contrast of an image is lowered when areas immediately surrounding the image are relatively dark (Figure 3.4). The figure shows two identical series of neutral squares, one series on a black background, the other on a white background. Note that each square on the black background appears to be lighter than its corresponding square on the white background. However, the effect is more evident for the darkest squares. As a result, the apparent luminance contrast—the perceived *difference* between the lightest and darkest squares—is diminished by the presence of the black background. A related effect, sometimes referred to as the *dark-surround effect*, occurs when an image such as a photographic slide, motion picture, or video image is projected in a darkened room and thus is surrounded by black.

Chromatic adaptation refers to adjustments of the visual mechanism in response to the average chromaticity of the stimulus (or collection of stimuli) to which the eyes are exposed. For example, when exposed sufficiently long to a reddish-yellow stimulus, such as tungsten light, the eye's longer-wavelength-sensitive receptors become somewhat desensitized and its shorter-wavelength-sensitive receptors become relatively more sensitive. (The white-balance adjustment of an electronic camera, which was

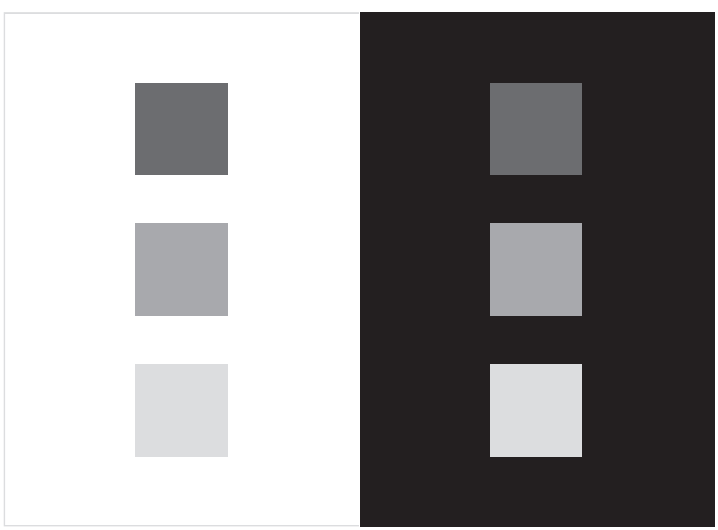


Figure 3.4 One manifestation of lateral-brightness adaptation. The three gray squares on the white background are physically identical to those on the black background, but the luminance contrast of the black-background series appears lower.

described earlier, is an approximate emulation of this process.) Chromatic adaptation helps the visual system interpret objects despite changes in the color of the illuminant. So a white flower generally will be recognized as white, regardless of the spectral composition of the illuminant under which it is viewed. Chromatic adaptation may not be *complete*, however, depending on the type of illumination that is used, the absolute level of that illumination, the extent to which the illumination fills the visual field, and certain other factors. For example, light from a dim tungsten lamp will continue to appear somewhat orange, even after the observer has had ample opportunity to adapt to it.

Colorimetry and human vision

From this discussion, it can be seen that the techniques of standard CIE colorimetry emulate the *first* stage of the human visual system, i.e., the image capture (trichromatic response) of the eye. Therefore colorimetry can be used to determine if two stimuli will produce the same trichromatic values, and this will predict whether those stimuli will visually match *if they are viewed under identical conditions.*

CIE colorimetry was developed specifically for that purpose. It is invaluable for determining metamerism and for specifying color-matching tolerances in a number of different industries. For example, it can be used to measure two paint samples to determine if they will match adequately under a specified illuminant. But CIE colorimetry does *not* emulate—nor was it designed to emulate—either the signal processing or the image-formation functions of the human visual system.

By itself, then, standard CIE colorimetry is *not* a direct predictor of *color appearance*. In Part II, it will be shown how that fact has very important consequences on the colorimetric design of color-imaging systems and media. Part III will show how the distinction between colorimetry and color appearance also has important consequences for digital color encoding, which is based on the measurement and numerical specification of color.

To emphasize the latter point, we will close Part I with some cautionary words from one of our mentors, E. J. Breneman. When he worked at Eastman Kodak Company, Breneman would attach the following warning to any CIE chromaticity diagram he encountered that was printed in color:

WARNING!

Odysseus had to go to great lengths to keep his sailors from being bewitched by the sirens. Similarly, people who want to specify color appearance with numbers must be wary of chromaticity diagrams with lovely colors on them. Such illustrations as this tend to imply (or the viewer tends to infer) that a particular perceived color is associated with a particular point on the diagram. This is false!

The chromaticity diagram is a stimulus specification. A given color perception can be elicited by a stimulus represented by just about any point on the chromaticity diagram, depending on the nature of the visual field and the viewer's state of adaptation.

Although a colored rendition of a chromaticity diagram does, perhaps, help to indicate to a novice that with a common illuminant, blue paint will have a chromaticity in the lower left, etc., that novice is likely to jump to all the wrong conclusions and consequently find the world of color very confusing.

In short, the uses of such illustrations probably should be confined to making lovely wall decorations in places that never will be visited by budding color scientists and engineers.

Summary of key issues

- In most imaging applications, a human observer, not a measuring instrument, is the ultimate judge of color quality.
- The human visual system performs the same three basic functions—image capture, signal processing, and image formation—performed by all other color-imaging systems.
- CIE colorimetry emulates only the first of these functions—trichromatic image capture.
- CIE colorimetry does not emulate the remaining functions—signal processing and image formation. Therefore, it does not measure color appearance.
- Color appearance is influenced by various psychophysical and psychological signal-processing effects.

PART II

The Nature of Color Images

Color images are the end product of color-imaging systems. In digital imaging systems, however, color images are also a source of input. In fact, in many systems, color images—and not original-scene objects—are the principal source of the color information that will be encoded. It is important to recognize that most images used for input are produced on color-imaging media and by devices typically designed for entirely different purposes.

For example, reflection images, such as those made using inkjet, thermal, conventional photographic, or other technologies, are designed for direct viewing under typical room conditions. Reflective media thus have colorimetric reproduction characteristics optimized for that purpose.

Similarly, transparency media such as photographic slide films and motion picture print films can be used as input sources on digital imaging systems. Yet these media are specifically designed to produce images that are projected, using particular types of projection lamps, in darkened rooms or theaters. Because they are optimized for that purpose, these media have colorimetric color-reproduction characteristics quite different from those of other types of imaging media.

Photographic negative films are an important input source for certain professional digital imaging applications, especially motion picture. Although these media actually are well suited for digital applications, few are designed specifically for that purpose. Most are instead designed to be optically printed onto another photographic medium. The colorimetric properties of negative films therefore are entirely different from those of any other imaging medium.

Electronic cameras are, of course, widely used as an input source for digital color-imaging systems. While they are well suited for digital applications, their colorimetric characteristics generally are designed to conform to standards originally developed for other industries, such as broadcast television. Moreover, since electronic cameras typically are used to record live scenes rather than reproduced images, their signals represent yet another fundamentally different type of input colorimetric information.

Digital Color Management: Encoding Solutions, Second Edition Edward J. Giorgianni and Thomas E. Madden
 ISBN: 978-0-470-51244-9

These basic differences among the various types of input sources greatly complicate the process of digitally encoding color. In order to determine the exact nature of those complications and to develop successful strategies for dealing with them, the fundamental colorimetric characteristics of each type of input source first must be examined.

In this part, the colorimetric characteristics of each of the basic types of imaging devices and media commonly used for input on digital color-imaging systems will be analyzed. The important consequences on digital color encoding that result from the unique colorimetric characteristics of each type of input source will be discussed. Understanding those consequences is essential for evaluating the capabilities and the limitations of existing color-encoding methods, which will be discussed later in Part III, and for determining the requirements for a comprehensive color-management system, which is the topic of Part IV.

4

Electronic Displays

Understanding the nature of color images requires knowledge both of the colorimetric properties of the images themselves and of the media and devices used to create them. In this chapter, the colorimetric characteristics of several important types of electronic devices used for image display will be examined. These devices will be described in terms of their basic technologies, intrinsic grayscale characteristics, and color primaries. We begin with a description of color CRT monitors.

CRT monitors—basic technology

A CRT-based monitor displays colors by producing various intensities of light from its red-, green-, and blue-light-emitting phosphors. RGB light intensities are determined by their respective red, green, and blue control signal voltages. In a typical computer monitor (Figure 4.1), these voltages are generated by digital-to-analog converters (DACs) from digital code values $R'G'B'$ sent to the DACs. (Note that the prime sign in this code-value designation is used to indicate values nonlinearly related to control signal voltages, as will be discussed in Chapter 5.)

CRT neutral characteristics

For any given set of monitor red, green, and blue phosphors, there is one and only one combination of *RGB* intensities that will produce a neutral (achromatic) stimulus having a given set of CIE *XYZ* tristimulus values. For example, Figure 4.2 shows the spectral power distributions for a representative set of CRT phosphor color primaries. Figure 4.3 illustrates how certain intensities of light from those phosphors can be combined to form a stimulus having a set of *XYZ* values equal to those of a reflective achromatic neutral patch illuminated by CIE Standard Illuminant D_{65}. The spectral power distribution of that illuminated reflective neutral is also shown in the figure.

Note that the monitor neutral and the illuminated reflective neutral stimuli are *highly* metameric, i.e., although they have the same *XYZ* tristimulus values, their spectral characteristics are *very* different. As a result, some observers—even those having "normal" color vision—will disagree as to whether neutrals displayed on the monitor perfectly match corresponding reflective neutrals.

Such disagreements should be expected when highly metameric stimuli are involved. Recall that CIE colorimetry allows for the prediction of visual matches *for a CIE Standard Colorimetric Observer*, but the color-matching functions of any particular observer are likely to differ somewhat from this standard. As a result, two stimuli having different spectral characteristics may appear identical to one observer but different to another. This is called *observer metamerism*.

Digital Color Management: Encoding Solutions, Second Edition Edward J. Giorgianni and Thomas E. Madden
 ISBN: 978-0-470-51244-9

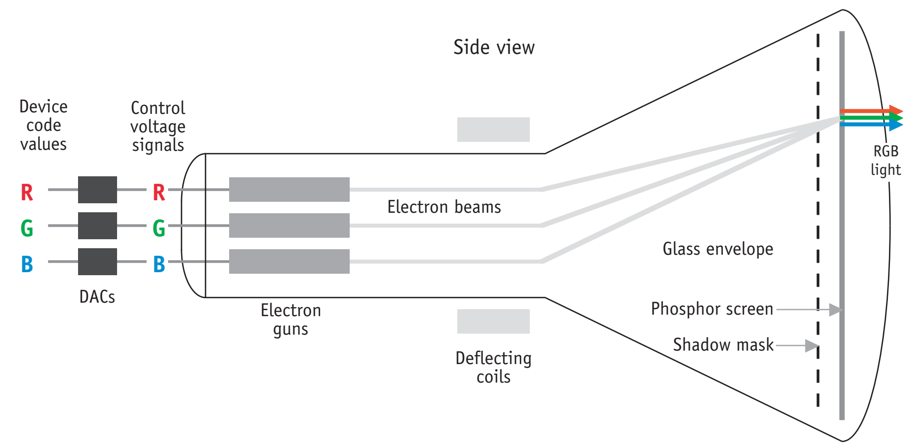


Figure 4.1 In a CRT-based color monitor, digital code values are converted to analog control-voltage signals. These voltage signals control the currents of electron beams that interact with phosphors on the CRT phosphor screen to generate light.

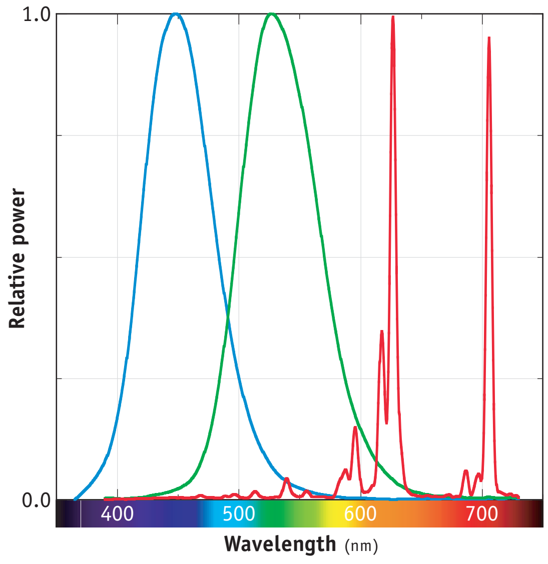


Figure 4.2 Spectral power distributions for light emitted by the red, green, and blue phosphors of a representative CRT-based display.

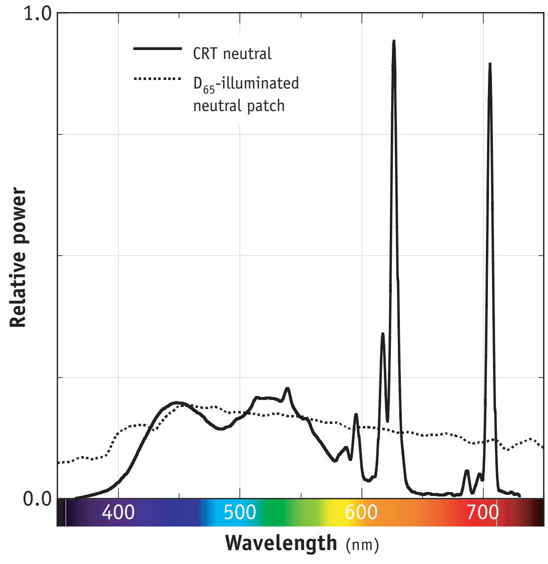


Figure 4.3 Relative spectral power distribution for a particular mixture of red, green, and blue light from a CRT. The *XYZ* values for this stimulus are identical to those of a corresponding achromatic neutral reflective patch illuminated by CIE Standard Illuminant D_{65}.

Observer metamerism can have important consequences in practical imaging systems. For example, a CRT-based display may be used to provide a softcopy preview of hardcopy images to be produced by a system, and a human operator may perform color-balance adjustments based on the CRT images. The operator's judgments will be influenced by the specific characteristics of his or her own visual system. If more than one operator makes these adjustments, the final output results may be quite inconsistent. This problem can be addressed by having each operator perform some type of visual matching to a reference hardcopy image. The results of that matching then can be factored into the adjustments subsequently made by each operator.

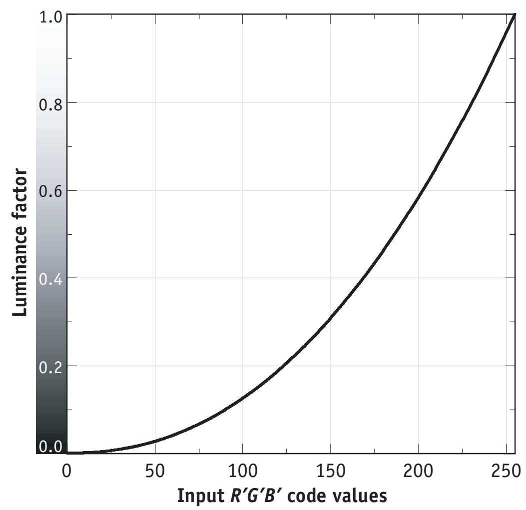


Figure 4.4 A CRT grayscale characteristic for a representative CRT-based display, expressed in terms of luminance factor versus equal $R'G'B'$ device code values.

CRT grayscale characteristics

A display device grayscale characteristic is a measure of how that device produces a series of neutrals, ranging from black to white, from a set of neutral (usually equal $R'G'B'$) input code values. In color-managed systems, a common practice is to set up display devices electronically such that the chromaticity of the stimulus produced from any set of equal $R'G'B'$ code values will not vary from the chromaticity of the white stimulus produced from a set of equal maximum $R'G'B'$ code values. If that condition is achieved, a display's grayscale can be characterized simply by measuring the luminance-factor values for a series of equal input $R'G'B'$ code values. The grayscale characteristic shown in Figure 4.4 is representative of many CRT-based computer monitors. This characteristic also is representative of broadcast studio monitors adjusted according to industry recommendations.

A CRT-based monitor grayscale characteristic generally can be described quite well by an equation in the following form:

$$Y = k\left(\frac{CV_{RGB}}{CV_{max}}\right)^{\gamma} + Y_0 \tag{4.1}$$

where Y is the luminance factor of the light emitted by the monitor CRT; CV_{RGB} is a set of equal $R'G'B'$ code values input to the monitor; Y_0 is an offset value, i.e., the luminance factor when the R', G', and B' code values equal zero; γ is the value of the exponent in the equation; and k is a normalizing factor determined such that $Y = 1.0$ when $CV_{RGB} = CV_{max}$, the maximum code value for the bit depth being used in the system.

This type of characteristic would be expected. CRTs inherently exhibit (more or less) a power-law relationship between output luminance and control voltage. The relationship between monitor code values and control voltages essentially is linear, so an approximate power-law relationship between output luminance levels and monitor code values should exist.

It generally is more appropriate to interpret grayscale characteristics if they are expressed in terms of logarithmic output values. That is because visual perceptions correlate more closely with logarithmic, rather than linear, luminance-factor values. For example, Figure 4.5a shows the grayscale characteristics, expressed in terms of *linear* luminance-factor values, for two monitors. The curves suggest that these grayscale characteristics are virtually identical.

Figure 4.5b shows the same two grayscale characteristics expressed in terms of *logarithmic* luminance-factor values. The logarithmic values

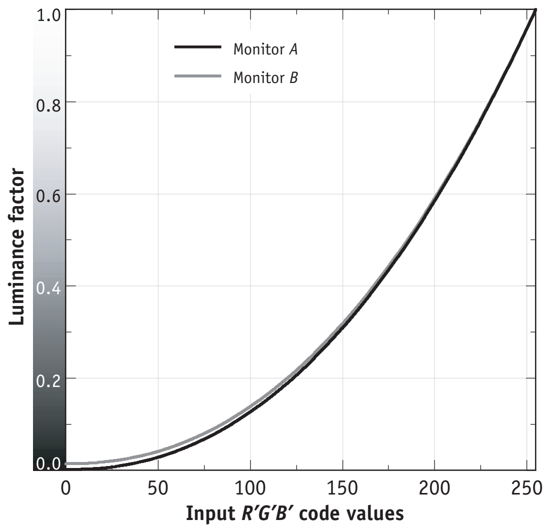


Figure 4.5a Grayscale characteristics for two CRT-based monitors, expressed in terms of luminance factor versus equal $R'G'B'$ device code values.

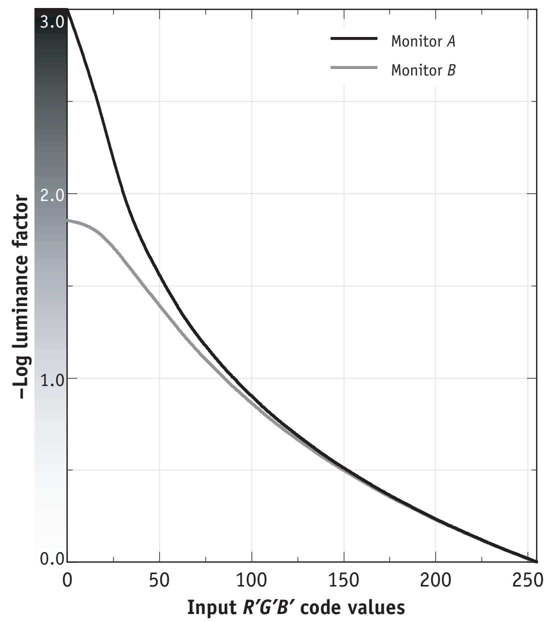


Figure 4.5b Grayscale characteristics for the two displays of Figure 4.5a, expressed in terms of negative log luminance factor versus equal $R'G'B'$ device code values. Logarithmic values provide a more meaningful representation of the significant visual differences in the grayscale characteristics of the two displays.

provide a much more meaningful description of the significant visual differences in the grayscale characteristics of the monitors. Note that the values shown in Figure 4.5b are *negative* logarithmic luminance-factor values, *relative to* the luminance-factor value produced by the maximum $R'G'B'$ input code values. Use of these values will allow electronic display grayscale characteristics to be compared later with those of hardcopy media, which typically are measured in terms of *visual density*. The comparison is straightforward because visual density is simply the negative logarithm of the luminance-factor value. (See Appendix B for more details.)

It should be noted here that some computer operating system software and some imaging applications load DAC lookup tables that alter the relationship between image code values and control voltages. Such tables might be used for any number of reasons. For example, a set of tables might help to produce more satisfactory monitor images from image data derived from a particular input device or medium. It is important to ensure that such tables are not present when attempting to measure the intrinsic characteristics of a display device itself.

CRT color characteristics

A color CRT-based monitor is an additive color device. It displays color by producing various intensities of light from its three primaries, i.e., from its red-, green-, and blue-light-emitting phosphors. Monitor primaries therefore are determined by the colorimetric properties of the particular phosphors used in the CRT.

An important property of an additive primary is that its chromaticity remains constant as its intensity level is varied. For example, Figure 4.6a shows an intensity series for the CRT primaries of Figure 4.2. As the figure illustrates, the wavelength-to-wavelength intensity ratios of each primary remain constant as a function of intensity level. Thus the chromaticity of each primary is independent of intensity level and remains constant throughout the series, as shown in Figure 4.6b. (Note that this property is *not* shared

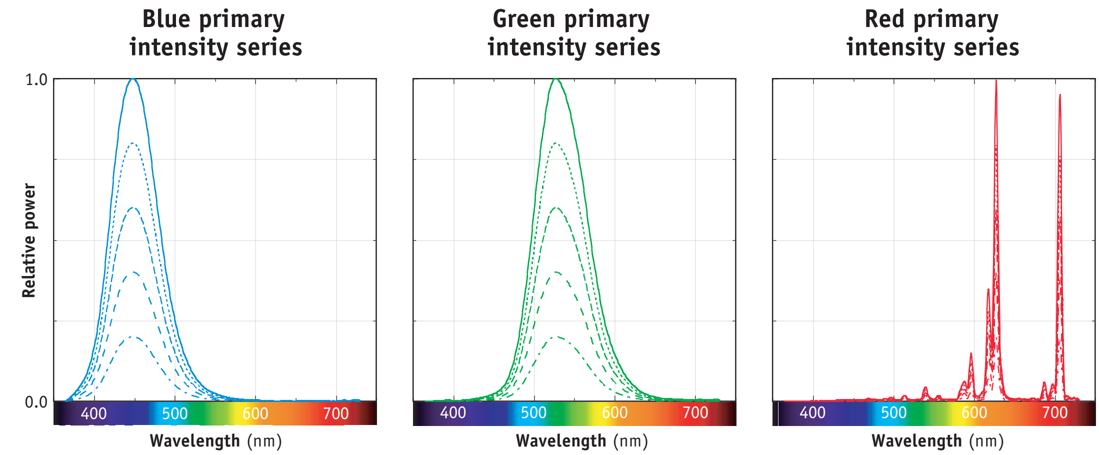


Figure 4.6a An intensity series for the CRT primaries of Figure 4.2.

by subtractive primaries such as dyes, inks, and pigments, as will be discussed in Chapter 6.)

When a set of three additive primaries is plotted on a chromaticity diagram, the primaries define vertices that form a triangle, the sides of which delineate the chromaticity boundaries of those primaries (Figure 4.7). Colors with chromaticities located outside this triangle cannot be produced by

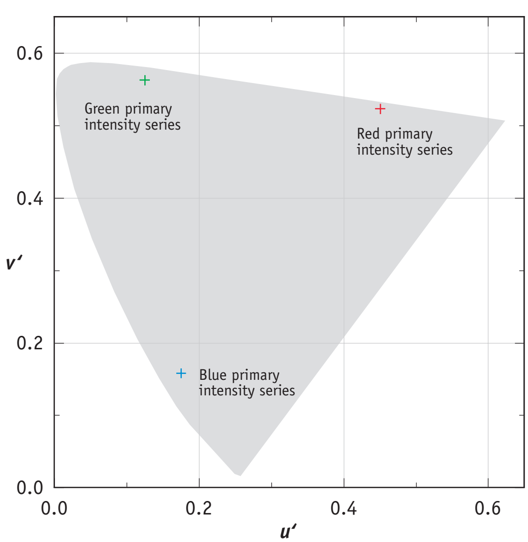


Figure 4.6b Chromaticities for the primary intensity series of Figure 4.6a. The wavelength-to-wavelength intensity ratios of each primary are independent of intensity level, therefore the chromaticity of each primary remains constant throughout the intensity series.

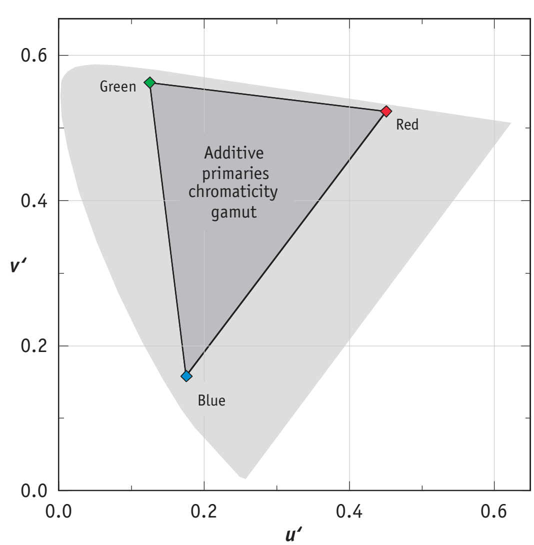


Figure 4.7 Chromaticity boundaries for a representative set of CRT primaries. Colors with chromaticities outside the boundaries cannot be produced by the primaries.

a device having primaries located at those vertices. Chromaticities located inside the triangle are achievable by the primaries. However, some colors having those chromaticities still may be unattainable due to limitations in the luminance levels achievable by the device at particular chromaticities.

Other electronic display technologies

In addition to CRT-based devices, electronic display devices based on other technologies also are used frequently in color-imaging systems. Such devices include LCD displays, plasma display panels (PDPs), field-emission displays (FEDs), organic light-emitting diode (OLED), and other types of electroluminescent displays, and Digital Light Processing® (DLP®) projectors. In each of these devices, input digital code values generate device drive signals that ultimately control the intensities of the display primaries. However, these devices can differ significantly in their intrinsic relationships between drive signals and output intensities and in the spectrophotometric and colorimetric characteristics of their color primaries.

In an LCD-based display, *RGB* device code values ultimately determine the intensities of displayed red, green, and blue light by controlling voltages applied across the liquid crystal layer in each pixel (Figure 4.8). The colorimetric characteristics of the display are determined by the light source used to back-light the display (typically a fluorescent or LED-based source) and the colored filters used on the individual pixels.

In a PDP-based display (Figure 4.9), *RGB* device code values determine the intensities of displayed red, green, and blue light by controlling the pulse rate of current flowing through the gas-filled cells of the display panel. The colorimetric characteristics of the display are determined for the most part by the phosphors used in the individual pixel cells. These can be the same phosphors used in typical color CRTs, in which case the two types of displays effectively would have the same color primaries. In other

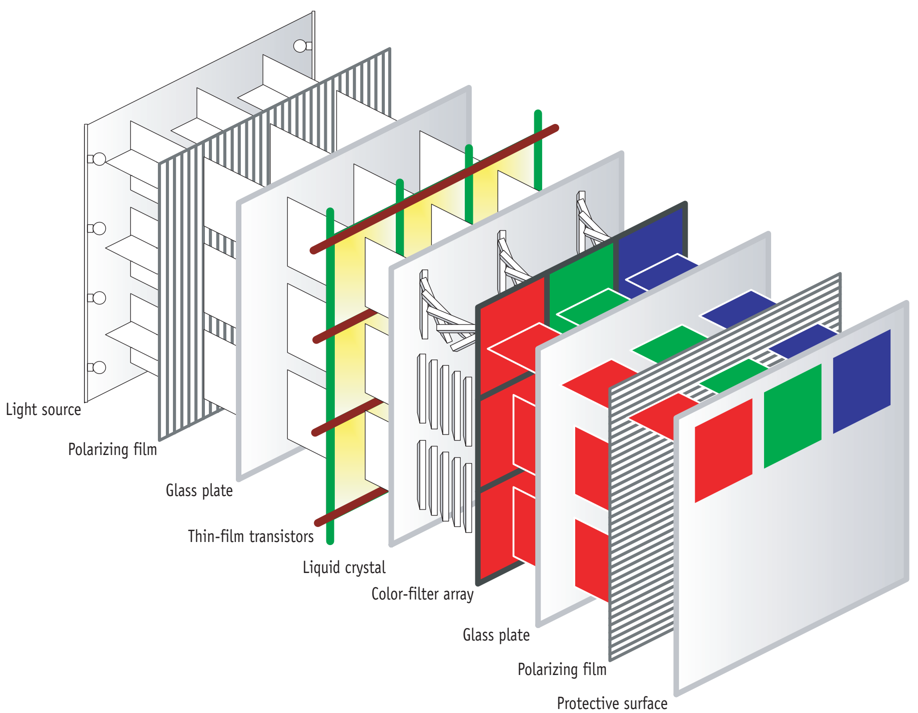


Figure 4.8 Basic technology of an LCD-based display.

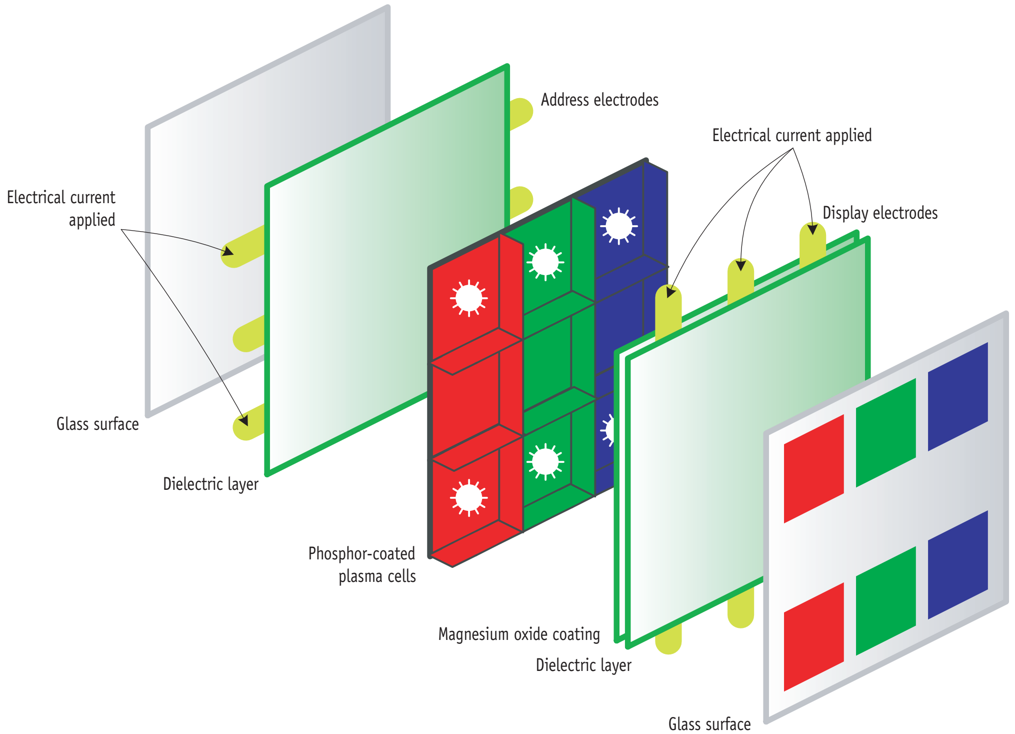


Figure 4.9 Basic technology of a PDP-based display.

cases, different phosphors might be used to increase the color gamut, luminance, or other properties of the PDP display. FED and SED (surface-conduction electron-emitter display) devices also use phosphor coatings as the light-emissive medium, and again their phosphors can be the same as those used in CRTs.

In an OLED-based display (Figure 4.10), *RGB* device code values control the amount of light emitted by special organic materials that emit red, green, blue, or white light. Active-matrix OLED displays comprise cathode, organic, and anode layers coated on a substrate that contains the electronic circuitry. Pixels are defined by the deposition pattern of the organic material. Each pixel is activated directly as a corresponding circuit delivers voltage to the cathode and anode materials, stimulating the middle light-emitting organic layer. Doping of the organic material helps determine the colorimetric characteristics of the display.

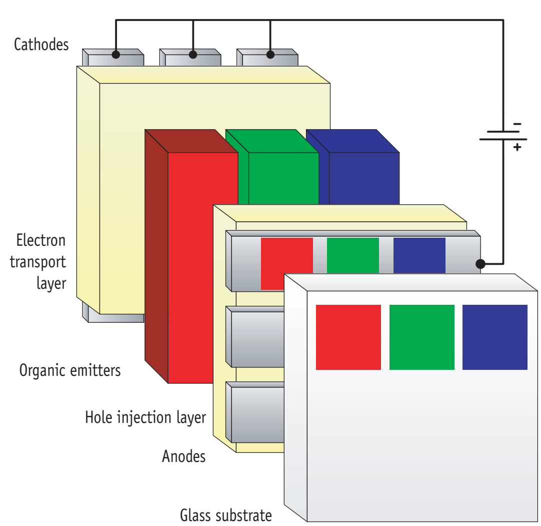


Figure 4.10 Basic technology of an OLED-based display.

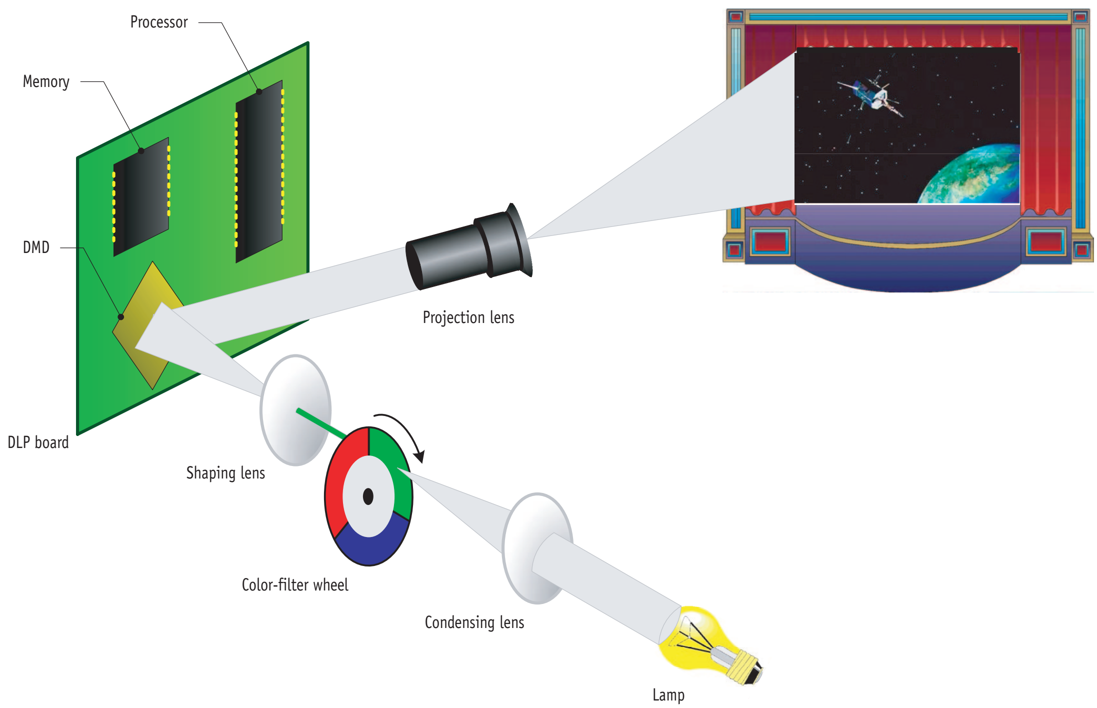


Figure 4.11 Basic technology of a single-chip DLP-based display.

In a DLP-based display, light intensities are modulated by a digital micromirror device (DMD)—a chip containing an array of microscopic reflectors that can be switched on and off up to several thousand times per second. Images are formed by projecting light reflected by the chip onto a front or back display surface. In a single-chip DLP device (Figure 4.11), colors are produced using a spinning color wheel situated between the device's light source and the DLP chip. The color wheel may be divided into red, green, and blue sectors, or it may include an additional transparent section or sections of other colors. The colorimetric characteristics of the display are determined principally by the light source and the colored filters. A three-chip DLP projector (Figure 4.12) uses a prism to split nominally white light from the light source into primary colors, each of which is routed to a separate DLP chip. The colorimetric characteristics of the display are determined principally by the light source, the color-separation prism, and color-trimming filters that may be placed between the prism and the DLP chips.

The diversity of these and other technologies used for electronic displays leads to a corresponding diversity in the colorimetric characteristics of the various types of display devices. Figures 4.13 and 4.14, for example, illustrate differences that can exist in the intrinsic grayscale characteristics and color primaries of several different types of device technologies. Figure 4.15 illustrates the degree of metamerism that can exist among the neutrals for these devices. In later chapters, we will develop strategies for dealing with such differences in color-managed imaging systems.

Determination of display colorimetry

In many color-management applications, it is important to be able to model, characterize, or otherwise predict the relationship between the input signal values for a display device and the colorimetry the device will produce. For example, if that relationship can be described by a mathematical model, display colorimetric values can be readily computed

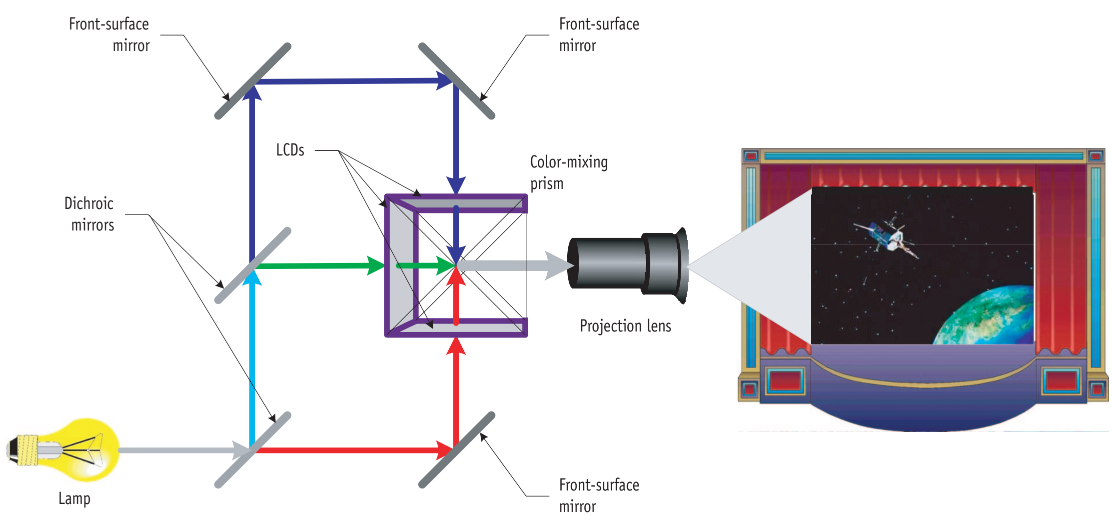


Figure 4.12 Basic technology of a three-chip DLP-based display.

from signal values input from sources such as digital cameras or computer workstations. Similarly, the signal values required to produce a desired set of colorimetric values on the display can be determined using an inverse of the model.

Figure 4.16 illustrates a model of an additive color-display device. In this model, digital code values first are transformed through the device characteristic curve to form normalized red-, green-,

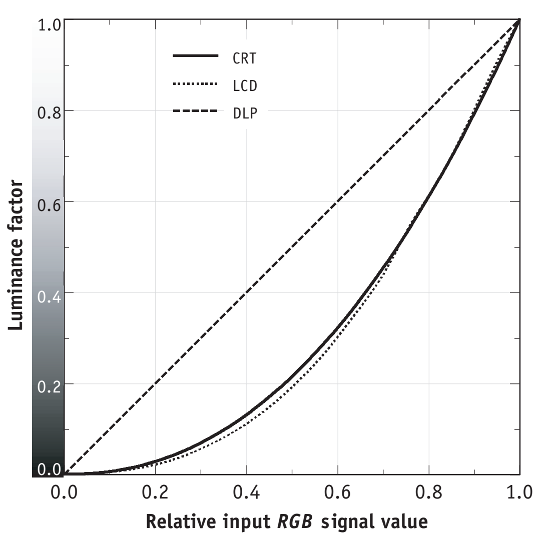


Figure 4.13 Comparison of the intrinsic grayscale characteristics of several different types of electronic displays.

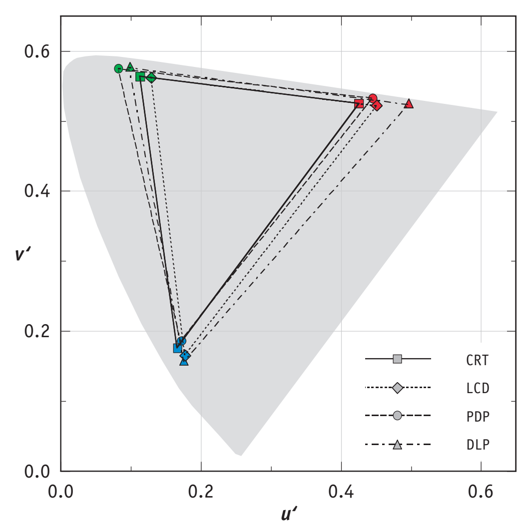


Figure 4.14 Comparison of the color primary chromaticities for representative samples of several different electronic display technologies. Differences can also exist among devices using the same basic technology.

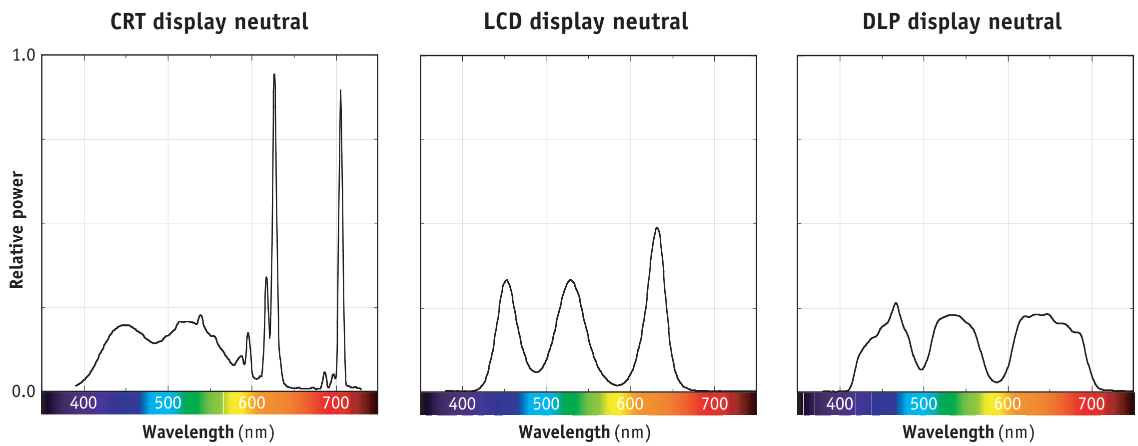


Figure 4.15 Comparison of the spectral properties of metameric neutrals for representative samples of several different types of electronic displays.

and blue-light-intensity values. Those *RGB* intensity values are then transformed through a primary-conversion matrix to form CIE *XYZ* tristimulus values. A procedure for determining primary-conversion matrices is described in Appendix H.

In some cases, simple models like this are quite adequate for describing the colorimetric behavior of a display device. Figure 4.17a, for example, shows

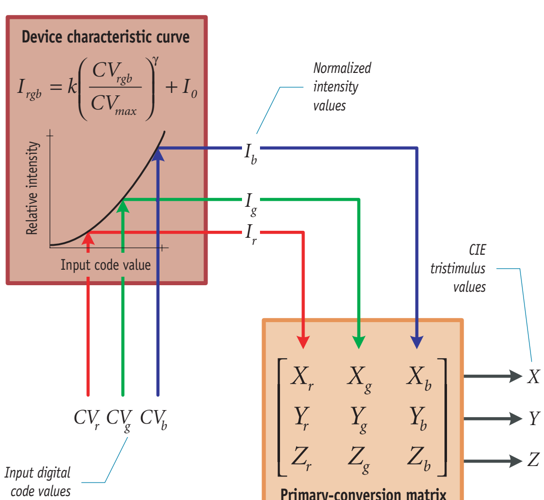


Figure 4.16 A simple model of an additive color-display device.

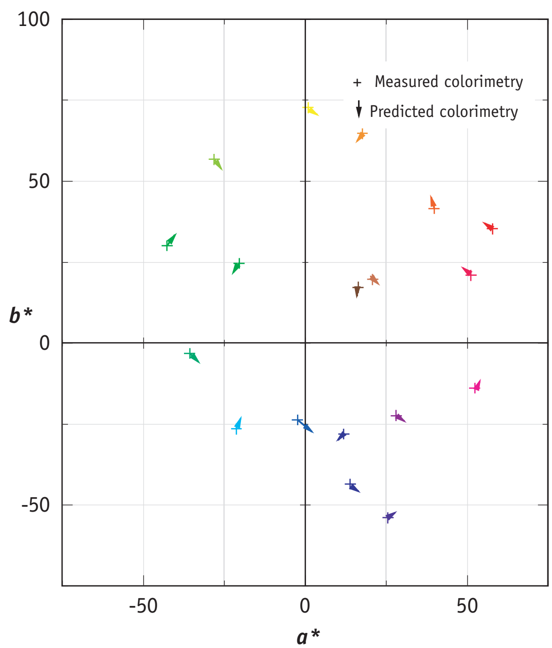


Figure 4.17a Correspondence of predicted and measured colorimetry for a particular CRT-based display where the predictions were made using the simple model shown in Figure 4.17b.

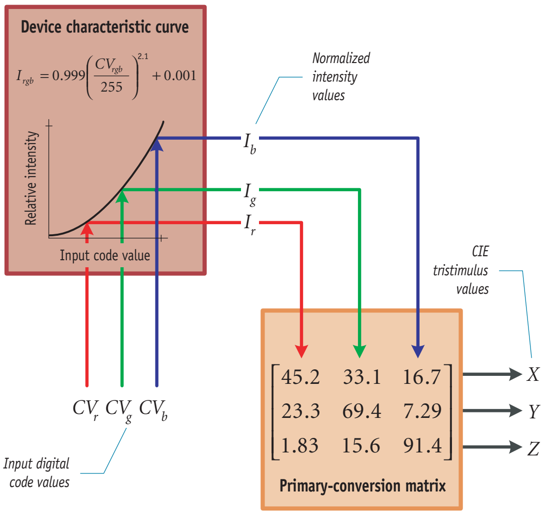


Figure 4.17b A simple model for a particular CRT-based display.

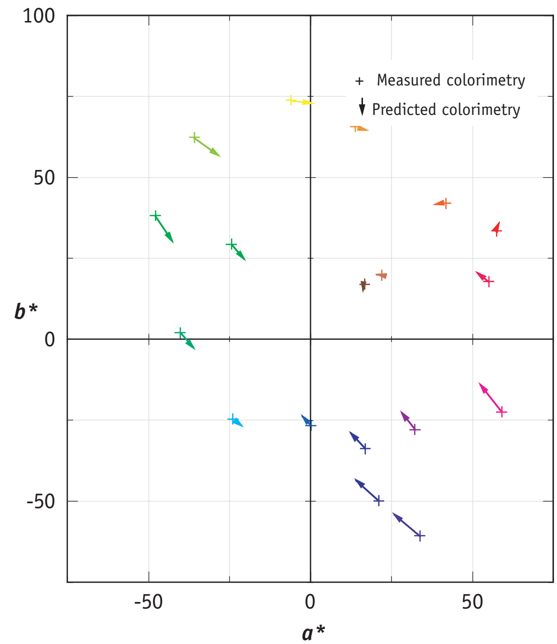


Figure 4.18a Correspondence of predicted and measured colorimetry for a more complex display device. The predictions were made using the simple model of Figure 4.16.

an excellent correspondence between predicted and measured colorimetry for a particular CRT-based display where the predictions were determined using the model shown in Figure 4.17b.

Often, however, such straightforward device models are not sufficient for accurate colorimetric predictions because display devices may additionally include various forms of analog or digital signal processing. If that signal processing is known, *and it is constant*, it can easily be included in the model of the device. For example, Figures 4.18a and 4.18b illustrate the predicted and actual colorimetry for a particular display device based on two different models. The first model, which uses only the grayscale characteristic and primaries of the display, produces the results shown in Figure 4.18a. The second model, which adds a control signal matrix and a set of lookup tables equivalent to those known to be operating within the device, produces the significantly more accurate results shown in Figure 4.18b.

In some cases, a device may have additional signal processing that is not readily apparent. For example, the analog portion of a display may include nonlinear signal amplification that effectively contributes one or more transformations, similar to a digital lookup table, to the signal processing. There also may be electronic crosstalk among the analog color signals, which essentially adds a color matrix to the processing of the signals. If the effects of such processing are constant and can be determined, perhaps from experimental measurements, they can be included in a device model to improve its accuracy.

Another complication, found in many types of electronic displays, is lack of spatial uniformity of light output. For example, when a uniform full-field signal is input to a display device, the light intensity at the edges of the displayed image should equal that at the image center. Often, however, the edges are darker than the center. Although that effect may not be objectionable or even noticeable in casual viewing, it may produce colorimetric deviations that are unacceptable for more critical applications. If the nonuniformity is constant, it can be emulated and included in the device model. Another approach, of course, would be to effectively eliminate

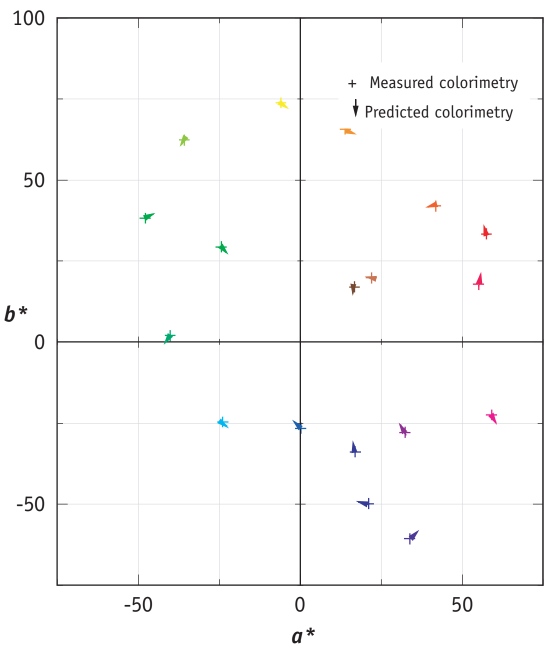


Figure 4.18b Correspondence of predicted and measured colorimetry for a more complex display device. The predictions were made using a model that includes a set of lookup tables and a matrix known to be incorporated in the signal processing of the device.

the nonuniformity by including appropriate spatial compensation in the processing of signals sent to the device.

In other cases, the prediction of display colorimetry can be far more problematic. It is important to remember that the vast majority of electronic display devices are designed to produce pleasing renditions of input images. For that reason, display devices may contain "adaptive" circuitry and algorithms that adjust the internal signal processing based on the content of the particular image to be displayed. In some of these devices, this adaptive signal processing can be disabled. When that cannot be done, the relationship between input signal values and colorimetric output varies according to the image content. Such devices are of little use in color-managed systems where colorimetric predictability is required.

The colorimetric behavior of some electronic displays also can be difficult or impossible to predict for other reasons. For example, a device may have an inadequate power supply. As a result, the relationship between a given set of input signal values and display light output will depend on the average signal level of the displayed image. This means that the device's intrinsic transfer function (signal in to light out) is not constant. Displays being considered for applications such as image preview and other critical color work should first be tested for this problem.

One simple test for this particular problem (sometimes referred to as *clamping failure*) consists of sequentially displaying a series of analysis images, such as those shown in Figure 4.19, and measuring the light output at the center test area. The digital code values of that area are identical in each image and are chosen to produce a mid-level gray. The test area is sufficiently large to prevent flare light from surrounding image areas to intrude. The surrounding area differs in each image in the series and ranges from full black to full white. The light output at the center area should, of course, measure identically for all of the displayed analysis images. On a good display, that indeed will be the case. This would indicate that there is a consistent and predictable relationship between input signal values and output light. On some displays, however, the center intensity of the displayed image will be affected by the average input image signal level, sometimes by as much as a factor of four! When the cause is an inadequate power supply, as the surround intensity increases the center area will darken due to increased draw on the power supply. Obviously, any attempt to predict the colorimetry of images on such devices is likely to be futile. Our experience is that such problems, when present, are inherent in brand-model designs and not in individual units. We have found no consistent correlation with device technology, price, or manufacturer. In our opinion, every model of every type of display device should be considered suspect until proven otherwise.

There may be other factors that complicate or prevent the accurate prediction of device colorimetry, including lack of color purity, sensitivity to image orientation, sensitivity to viewing angle, and internal flare. Such problems may not always be evident when viewing typical pictorial images, and they might go

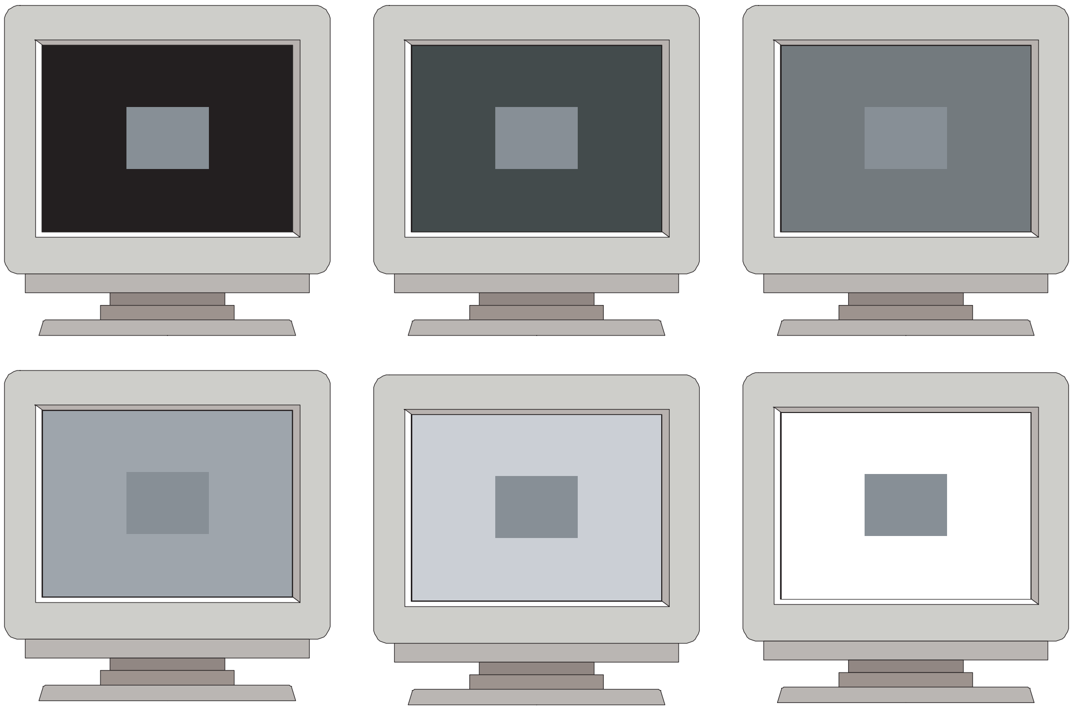


Figure 4.19 A series of analytical images for evaluating electronic display devices. Colorimetric measurements made of the center area should not vary as the surrounding area is changed.

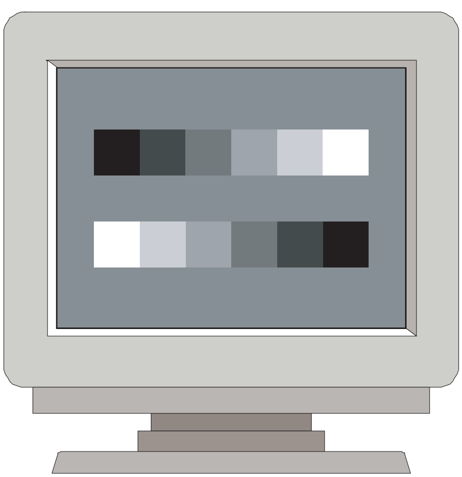


Figure 4.20 A test image for evaluating electronic display devices. Colorimetric measurements made of the grayscales should be identical for the two orientations.

undetected for some time. Therefore, it is good practice to evaluate display devices using test images appropriate for uncovering particular problems. For example, an image of opposing but otherwise identical grayscales, like that shown in Figure 4.20, often exposes problems in scanning-type output devices where the output level in a given area can be affected by the levels of preceding areas. It is also good practice to test the robustness of colorimetric predictions using a variety of verification-test images that are independent of any test images used to derive a model or device characterization, and in which a variety of different test colors, different geometric layouts, and different backgrounds are included.

Summary of key issues

- Virtually all electronic display devices are based on additive color technology.

- Different types of electronic display devices may have significantly different intrinsic grayscale characteristics.
- Different types of electronic display devices may have significantly different color primaries.
- Color matches involving electronic displays and hardcopy images generally are highly metameric.
- Color matches among different types of electronic displays also may be highly metameric.
- An electronic display device can be modeled or characterized accurately if the relationships between input code values and output colorimetry are consistent.
- Some electronic display devices cannot be accurately modeled or characterized because the relationships between input code values and output colorimetry are inconsistent due to various electronic and optical effects.

5

Electronic Imaging Systems

In Chapter 4, the colorimetric properties of several different types of electronic displays were described. Forming complete imaging systems based on such displays additionally requires an image-capture device and a means for applying appropriate image signal processing. In this chapter a complete electronic imaging system will be developed, and optimum image-capture characteristics and signal processing will be determined.

The example system will be analyzed in terms of its basic imaging functions (Figure 5.1). The system will consist of a digital still camera, which will perform the functions of image capture and signal processing, and a CRT-based monitor for image display. Why use a CRT display in this analysis? Although they now are being replaced by other types of electronic display devices, CRTs were virtually universal when industry standards were developed for conventional video and high-definition television (HDTV) cameras, broadcast television signals, and many digital image-encoding metrics, all of which continue to be in wide use. To conform to these existing standards, newer electronic cameras still must be capable of producing CRT-compatible color-image signals, whether or not the images they produce are ever reproduced on a CRT display.

Similarly, newer display devices must be capable of working with color signals originally intended for CRT displays. So although CRT technology eventually may become obsolete, its influence on electronic imaging components and systems will continue for some time. It also should be noted that the basic analytical approach described in this chapter can be applied to any electronic imaging system using any type of additive color-display technology.

The example system will be examined using an original scene consisting of a test target composed of six spectrally nonselective neutral patches and six saturated color patches. Figure 5.2a shows the general layout of the target and the spectral reflectance characteristics of each of the test patches. The target will be illuminated by a reference light source simulating CIE Standard Illuminant D_{65} (Figure 5.2b). Colorimetry for the target colors will be computed from the spectral reflectances of the patches and the spectral power distribution of the light source. Colorimetry of the color stimuli reproduced on the monitor will be measured and compared to that of the original color stimuli of the illuminated test target (Figure 5.3).

To get started, let us assume *for the moment* that the objective is for the system to produce a perfect colorimetric (metameric) match between the original illuminated test target and its reproduction on the monitor. This means that the *XYZ* tristimulus values measured from the displayed image should equal those computed for the D_{65}-illuminated target.

Digital Color Management: Encoding Solutions, Second Edition Edward J. Giorgianni and Thomas E. Madden
 ISBN: 978-0-470-51244-9

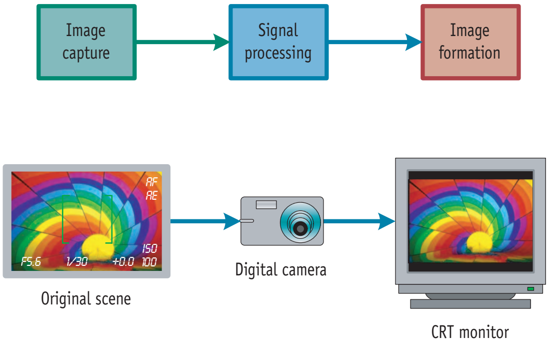


Figure 5.1 The basic imaging functions of image capture, signal processing, and image formation in an example electronic imaging system.

Monitor setup

In the analysis of this example system, a colorimeter will be used to measure the CIE *XYZ* values of monitor stimuli. Measurements will be made with the colorimeter and monitor arranged such that the measurement is *flareless*, i.e., the only light measured is that produced by the monitor itself, free of any stray room light.

Because the test target in the example system is to be illuminated by D_{65}, the monitor white also is set to the chromaticity of D_{65}. This means that the monitor is adjusted electronically such that a white having chromaticity coordinates equal to those of D_{65} ($x = 0.3127$, $y = 0.3290$) is produced when maximum $R'G'B'$ input code values are used

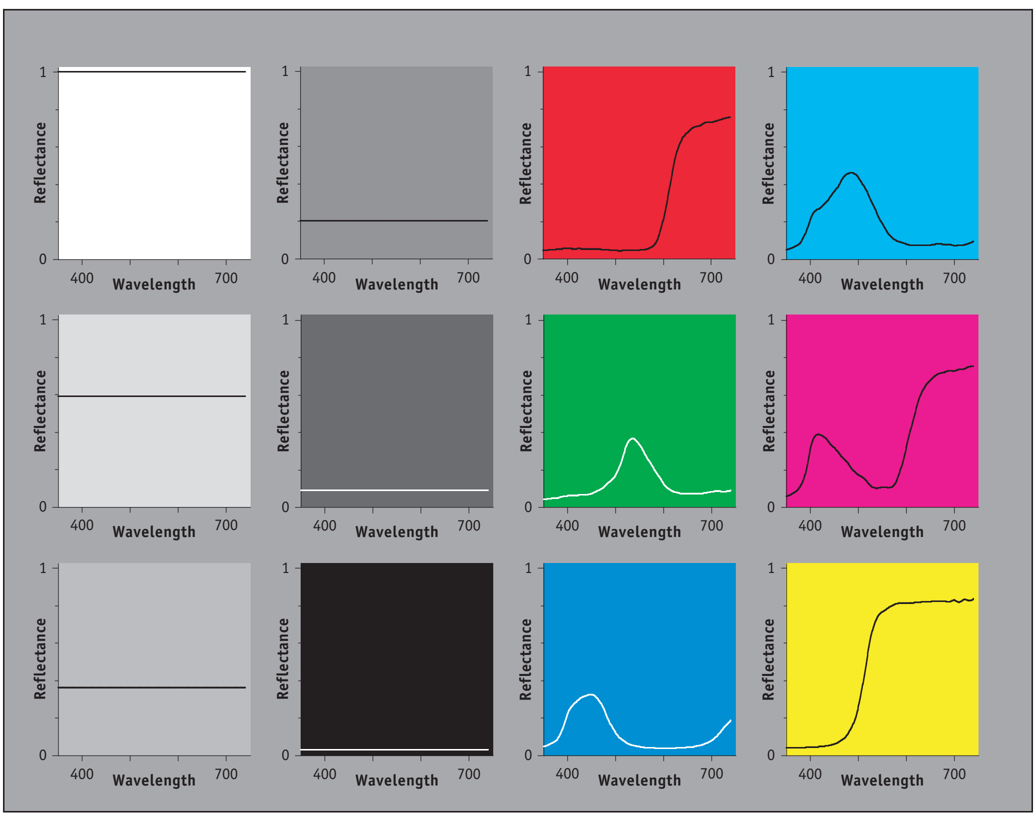


Figure 5.2a General layout and spectral reflectances for the patches of the example test target.

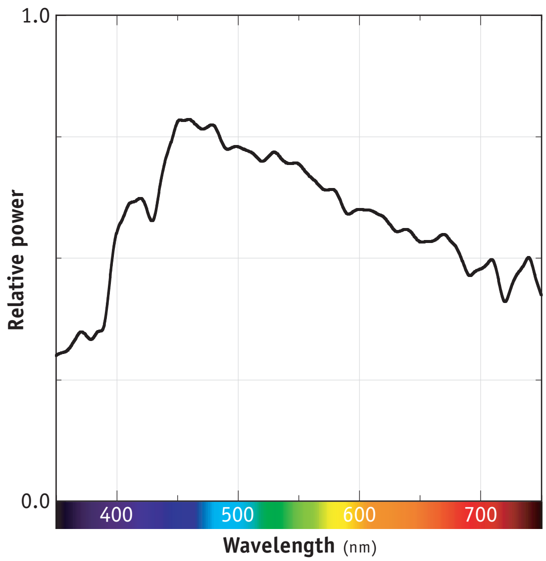


Figure 5.2b Spectral power distribution for CIE Standard Illuminant D_{65}.

($R' = 255$, $G' = 255$, $B' = 255$ in this 8-bit-per-channel system). In addition, the absolute luminance of the monitor white is adjusted to equal the absolute luminance of the illuminated perfect white (the first test patch) of the test target. Finally, the monitor is adjusted such that its grayscale "tracks"; that is, the chromaticity of the light it emits remains constant (equal to the chromaticity of D_{65}) for any set of equal input $R'G'B'$ code values. Because its chromaticity does not vary, the grayscale can be characterized simply by measuring the monitor luminance-factor values for a series of equal input $R'G'B'$ code values.

The grayscale characteristic of the particular monitor used in the example system, expressed in terms of luminance versus input equal $R'G'B'$ code value, is shown in Figure 5.4a. The characteristic is shown in terms of negative log luminance-factor versus input equal $R'G'B'$ code value in Figure 5.4b. The grayscale characteristic is described well by Equation (4.1), as in Chapter 4 and shown again

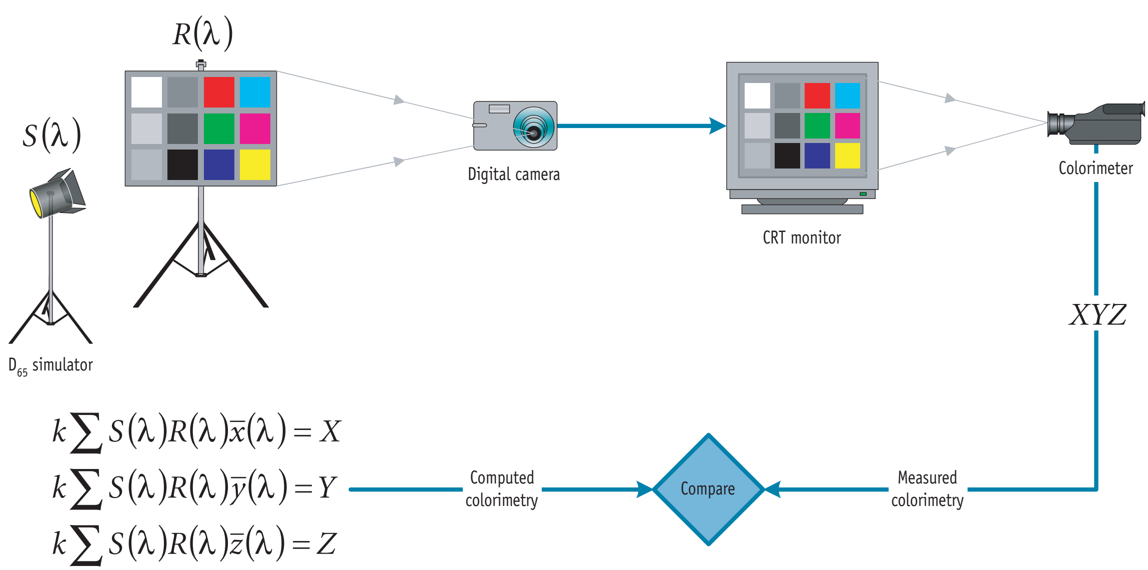


Figure 5.3 Arrangement of the test target, digital still camera, monitor, and colorimeter, as described in the text. In this example system, the measured colorimetry of the monitor image is compared to the computed colorimetry of the illuminated test target.

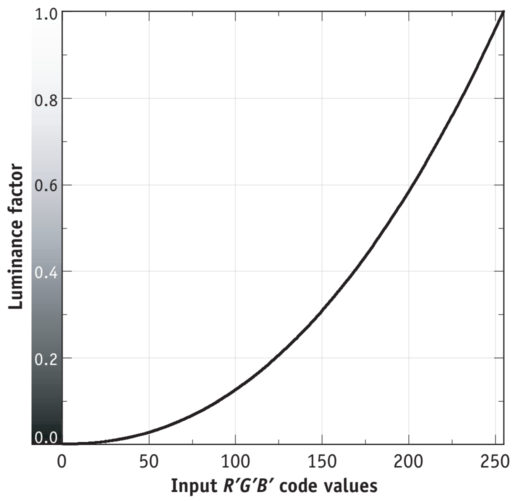


Figure 5.4a The CRT grayscale characteristic for the monitor used in the example system, expressed in terms of luminance factor versus equal input $R'G'B'$ values.

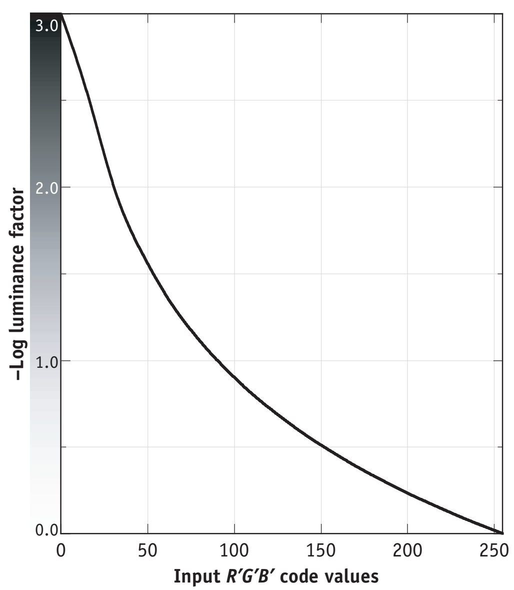


Figure 5.4b The CRT grayscale characteristic for the monitor used in the example system, expressed in terms of negative log luminance factor versus equal input $R'G'B'$ values.

here as Equation (5.1a). For the example monitor, the value of the normalizing factor, k, is 0.999; the value of the exponent, γ, is 2.22; and the value of the offset, Y_0, is 0.001, as shown in Equation (5.1b):

$$Y = k\left(\frac{CV_{RGB}}{CV_{max}}\right)^{\gamma} + Y_0 \tag{5.1a}$$

$$Y = 0.999\left(\frac{C_{R'G'B'}}{255}\right)^{2.22} + 0.001 \tag{5.1b}$$

Camera grayscale characteristics

The grayscale characteristic of a digital camera defines the $R'G'B'$ code values output by the camera in response to a series of neutral stimuli, i.e., stimuli having the chromaticity of the reference light source. The camera of the example system has been adjusted such that its grayscale characteristic "tracks"; that is, it produces equal $R'G'B'$ code values for neutral stimuli of any given luminance. The camera's grayscale characteristic therefore can be expressed in terms of equal $R'G'B'$ code values as a function of scene luminance-factor value (or scene exposure-factor value).

The appropriate characteristic for the camera grayscale can be determined based on the system's monitor grayscale characteristic. Since, for the moment, the goal for the example system is to have the colorimetry reproduced on the monitor equal that of the original illuminated test target, the grayscale characteristic of the digital still camera must be the *mathematical inverse* of the monitor characteristic. In other words, the characteristic must be such that when the camera is used together with the monitor, a one-to-one *overall system* grayscale characteristic—from the original illuminated test target to the final monitor reproduction—will be produced.

A camera grayscale characteristic that is the inverse of the monitor grayscale characteristic can be achieved by the use of appropriate camera signal processing. That signal processing often is referred to as *gamma correction* because it may include a power function with an exponent that is the reciprocal of the monitor gamma. In the example system, the monitor gamma is 2.22. Therefore, the exponent of the power function for the camera grayscale characteristic would be 1/2.22, or 0.45.

The actual inverse digital camera grayscale characteristic for the particular monitor used in this example, and which was described by Equation (5.1), is as follows:

$$C_{R'G'B'} = 255\left(\frac{Y-0.001}{0.999}\right)^{0.45} \tag{5.2}$$

where $C_{R'G'B'}$ is a set of equal $R'G'B'$ camera output code values and Y is the luminance factor of the neutral stimulus being captured.

A plot of this inverse digital camera characteristic is shown in Figure 5.5a. The characteristic is expressed in terms of equal output $R'G'B'$ code values as a function of scene luminance factor. In Figure 5.5b, the same camera characteristic is expressed in terms of equal output $R'G'B'$ code values as a function of *log* scene luminance factor.

Use of a camera having this particular characteristic, together with the monitor of the example system, will result in a one-to-one overall system grayscale characteristic. The basic steps involved in the formation of a system grayscale from camera and monitor grayscales are illustrated in Figure 5.6a. A one-to-one system grayscale is shown in terms of linear values in Figure 5.6b and shown again in terms of logarithmic values in Figure 5.6c.

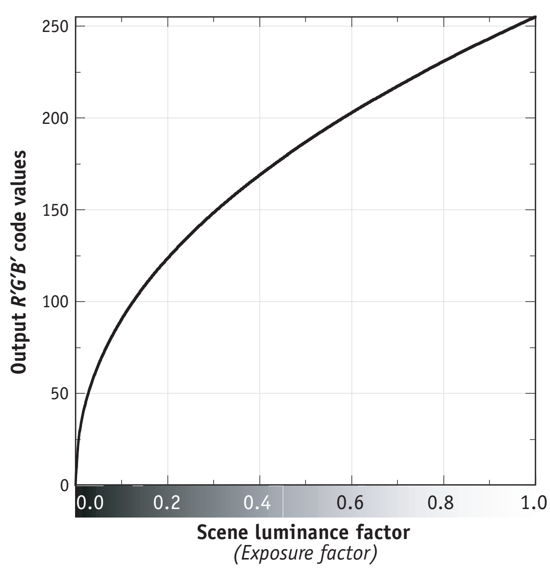


Figure 5.5a Monitor-inverse digital camera characteristic, expressed in terms of equal output $R'G'B'$ code values as a function of linear scene luminance factor.

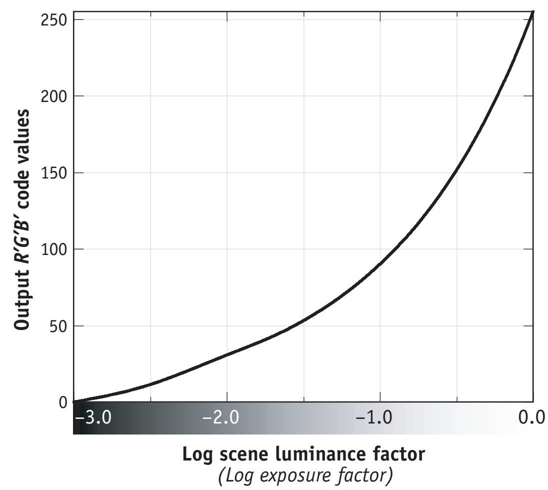


Figure 5.5b Monitor-inverse digital camera characteristic, expressed in terms of equal output $R'G'B'$ code values as a function of log scene luminance factor.

The achievement of a one-to-one system grayscale reproduction is an important and necessary step toward achieving an overall colorimetric color-reproduction system. This system grayscale ensures that the measured output luminance-factor values of neutrals will equal the measured input luminance-factor values for those same neutrals. In addition, the camera and monitor electronic setups described earlier ensure that the chromaticity values of the output neutrals will equal those of the corresponding input neutrals.

However, these conditions alone will not necessarily result in the chromaticities of output *non-neutral* colors being equal to the chromaticities of the corresponding input non-neutral colors. In order to also achieve that objective, appropriate red, green, and blue spectral responsivities for the

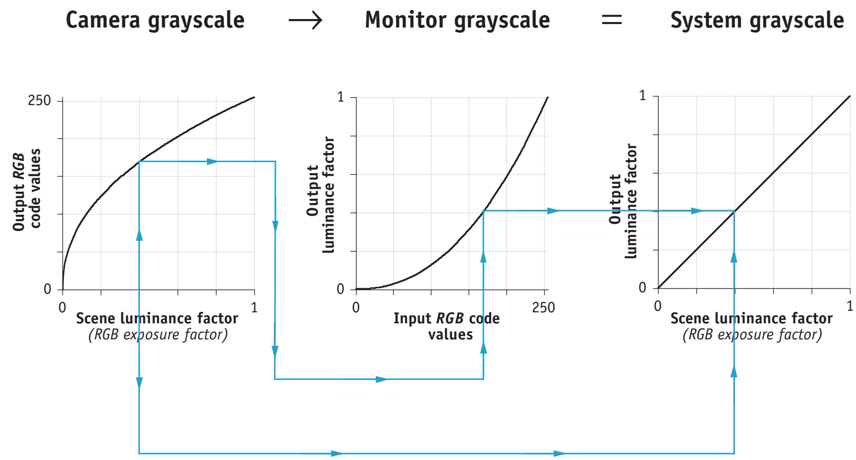


Figure 5.6a Formation of an overall system grayscale. Scene luminances produce camera output code values. When these code values are input to a monitor, output luminances are produced.

electronic camera first must be determined. Moreover, any additional required color signal processing must be defined for the overall system. To do this, it is necessary to further investigate the color characteristics of the system monitor.

Monitor color characteristics

Color-matching functions for an additive color display can be computed from the color-matching functions of the CIE Standard Colorimetric Observer

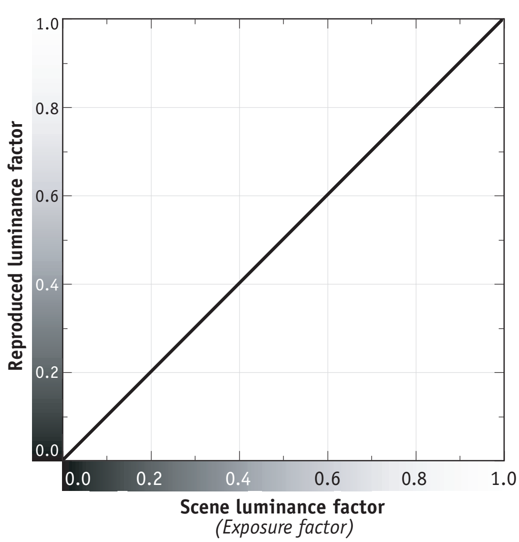


Figure 5.6b A one-to-one overall system grayscale characteristic, expressed in terms of reproduced luminance factor as a function of original-scene luminance factor.

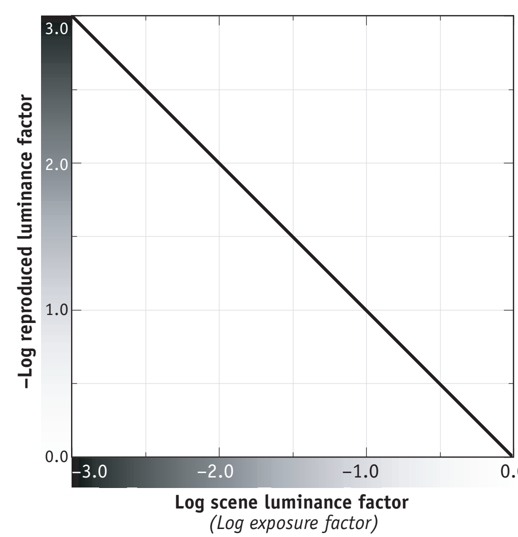


Figure 5.6c A one-to-one overall system grayscale characteristic, expressed in terms of negative log reproduced luminance factor as a function of log original-scene luminance factor.

and the chromaticity coordinates (x, y) of the display's red, green, and blue color primaries. (See Appendix I for details.) The chromaticities of the primaries of the monitor used in the example system (Figures 5.7a and 5.7b) would generate the set of color-matching functions shown in Figure 5.8.

The resulting monitor color-matching functions can be used to calculate the relative intensities of red, green, and blue light that the monitor must generate in order to produce a stimulus that colorimetrically matches another given color stimulus. Table 5.1 shows the relative red, green, and blue monitor-intensity values needed to match each illuminated patch of the test target used in this example. The equations used to calculate those values are as follows:

$$
\begin{aligned}
I_r &= k_{m_r} \sum_{\lambda} S(\lambda) R(\lambda)\, \bar{r}_m(\lambda) \\
I_g &= k_{m_g} \sum_{\lambda} S(\lambda) R(\lambda)\, \bar{g}_m(\lambda) \qquad (5.3) \\
I_b &= k_{m_b} \sum_{\lambda} S(\lambda) R(\lambda)\, \bar{b}_m(\lambda)
\end{aligned}
$$

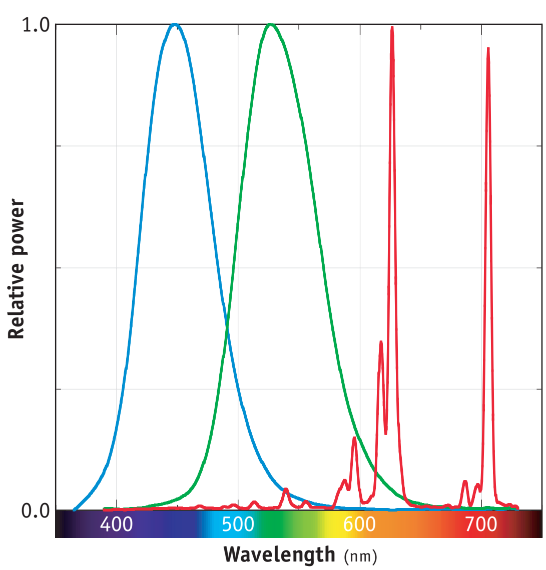


Figure 5.7a Spectral power distributions for light produced by the red-, green-, and blue-light-emitting CRT phosphors of the monitor used in the example system.

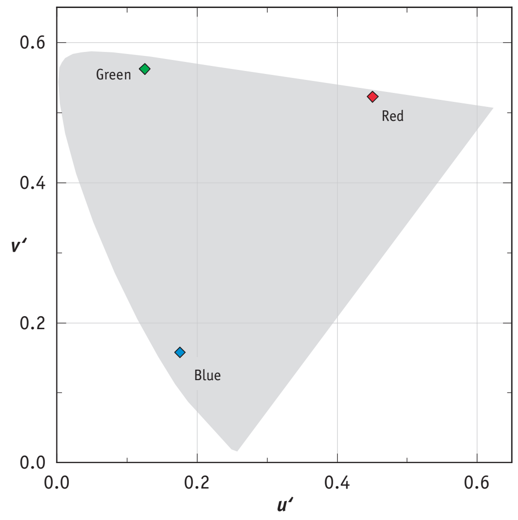


Figure 5.7b Chromaticities for light produced by the red-, green-, and blue-light-emitting CRT phosphors of the monitor used in the example system.

where I_r, I_g, and I_b are the red, green, and blue relative intensity values; $S(\lambda)$ is the relative spectral power distribution of the light source; $R(\lambda)$ is the spectral reflectance of the test patch; $\bar{r}_m(\lambda)$, $\bar{g}_m(\lambda)$, and

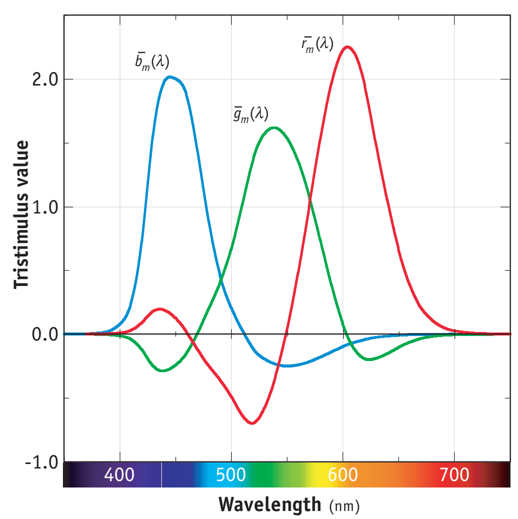


Figure 5.8 Color-matching functions for the red, green, and blue primaries of the monitor used in the example system.

Table 5.1 Monitor *RGB* relative intensity values, I_R, I_G, I_B, required to colorimetrically match the illuminated test target.

Test-Target Color	Intensities Required for Colorimetric Matching		
	I_R	I_G	I_B
Neutral 1	1.0000	1.0000	1.0000
Neutral 2	0.5890	0.5890	0.5890
Neutral 3	0.3630	0.3630	0.3630
Neutral 4	0.2000	0.2000	0.2000
Neutral 5	0.0890	0.0890	0.0890
Neutral 6	0.0320	0.0320	0.0320
Red	0.3953	0.0156	0.0444
Green	0.0893	0.3209	0.0527
Blue	0.0313	0.0407	0.3322
Cyan	−0.0010	0.2631	0.4038
Magenta	0.4598	0.0748	0.3290
Yellow	0.9037	0.6827	−0.0400

$\overline{b}_m(\lambda)$ are the red, green, and blue color-matching functions for the monitor primaries; and k_{m_r}, k_{m_g}, and k_{m_b} are normalizing factors determined such that I_r, I_g, and $I_b = 1.00$ for the first test patch (a perfect white reflector).

For colorimetric matching, the camera red, green, and blue signals for any test stimulus ultimately must result in the monitor red-, green-, and blue-light intensities required to produce a stimulus having colorimetry identical to that of the test stimulus. Since signal processing already has been incorporated such that a one-to-one relationship of camera exposure-factor values and monitor intensity values has been achieved for neutral stimuli, all that remains is to determine the particular camera spectral responsivities that also will produce the exposure-factor values required for the colorimetric matching of non-neutral stimuli.

The spectral responsivities required to do that can be determined by setting the camera *RGB* exposure equations, Equations (2.1), equal to the monitor *RGB* intensity equations, Equations (5.3):

$$\begin{aligned} R_{exp} &= I_r \\ G_{exp} &= I_g \\ B_{exp} &= I_b \end{aligned} \tag{5.4a}$$

or

$$\begin{aligned} k_{c_r} \sum_{\lambda} S(\lambda)\, R(\lambda)\, r_c(\lambda) &= k_{m_r} \sum_{\lambda} S(\lambda)\, R(\lambda)\, \overline{r}_m(\lambda) \\ k_{c_g} \sum_{\lambda} S(\lambda)\, R(\lambda)\, g_c(\lambda) &= k_{m_g} \sum_{\lambda} S(\lambda)\, R(\lambda)\, \overline{g}_m(\lambda) \\ k_{c_b} \sum_{\lambda} S(\lambda)\, R(\lambda)\, b_c(\lambda) &= k_{m_b} \sum_{\lambda} S(\lambda)\, R(\lambda)\, \overline{b}_m(\lambda) \end{aligned} \tag{5.4b}$$

which reduces to

$$\begin{aligned} r_c(\lambda) &= \frac{k_{m_r}}{k_{c_r}} \overline{r}_m(\lambda) \\ g_c(\lambda) &= \frac{k_{m_g}}{k_{c_g}} \overline{g}_m(\lambda) \\ b_c(\lambda) &= \frac{k_{m_b}}{k_{c_b}} \overline{b}_m(\lambda) \end{aligned} \tag{5.5}$$

What this means, in general terms, is that colorimetric matching between the original and reproduction can be achieved if the red, green, and blue spectral responsivities, $r_c(\lambda)$, $g_c(\lambda)$, and $b_c(\lambda)$ of the image-capture device are *proportional* to the respective red, green, and blue color-matching functions $\overline{r}_m(\lambda)$, $\overline{g}_m(\lambda)$, and $\overline{b}_m(\lambda)$ for the display

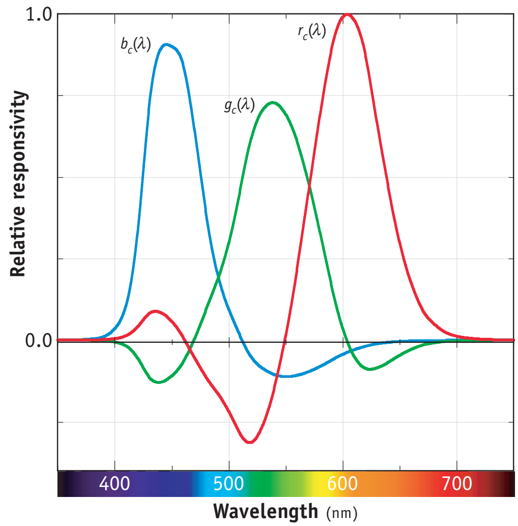


Figure 5.9 Ideal spectral responsivities for a digital still camera to be used with a monitor having the primaries of Figures 5.7a and 5.7b. The shapes of these responsivities equal the respective shapes of the monitor color-matching functions shown in Figure 5.8.

device. The ratios k_{m_r}/k_{c_r}, k_{m_g}/k_{c_g}, and k_{m_b}/k_{c_b} define the appropriate proportionalities. A simpler way of saying this is that the red spectral responsivity curve for the image-capture device should have the same "shape" as the red color-matching-function curve for the output device, and likewise for the green and blue channels.

The digital still camera for the example system, then, should have relative red, green, and blue spectral responsivities that correspond directly to the red, green, and blue color-matching functions for the monitor (Figure 5.8). The ideal spectral responsivities, white balanced for D_{65}, are shown in Figure 5.9. A camera having these responsivities would generate the *RGB* exposure-factor values shown in Table 5.2. Note that those values are *identical* to the required monitor *RGB* relative intensity values that were shown previously in Table 5.1.

It would appear that achieving colorimetric color reproduction is quite simple. All that seems necessary is a one-to-one system grayscale characteristic and a camera having spectral responsivities like those just defined. However, if one were to attempt to *construct* a camera having such responsivities, a serious

Table 5.2 The *RGB* exposure-factor values, R_{exp}, G_{exp}, B_{exp}, for a digital camera having ideal spectral responsivities, equivalent to the color-matching functions of the monitor *RGB* primaries, exactly equal the monitor *RGB* relative intensity values, I_R, I_G, I_B, required for colorimetric matching.

	Intensities Required for Colorimetric Matching			Exposures from Ideal Camera Spectral Responsivites		
Test-Target Color	I_R	I_G	I_B	R_{ideal}	G_{ideal}	B_{ideal}
Neutral 1	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
Neutral 2	0.5890	0.5890	0.5890	0.5890	0.5890	0.5890
Neutral 3	0.3630	0.3630	0.3630	0.3630	0.3630	0.3630
Neutral 4	0.2000	0.2000	0.2000	0.2000	0.2000	0.2000
Neutral 5	0.0890	0.0890	0.0890	0.0890	0.0890	0.0890
Neutral 6	0.0320	0.0320	0.0320	0.0320	0.0320	0.0320
Red	0.3953	0.0156	0.0444	0.3953	0.0156	0.0444
Green	0.0893	0.3209	0.0527	0.0893	0.3209	0.0527
Blue	0.0313	0.0407	0.3322	0.0313	0.0407	0.3322
Cyan	−0.0010	0.2631	0.4038	−0.0010	0.2631	0.4038
Magenta	0.4598	0.0748	0.3290	0.4598	0.0748	0.3290
Yellow	0.9037	0.6827	−0.0400	0.9037	0.6827	−0.0400

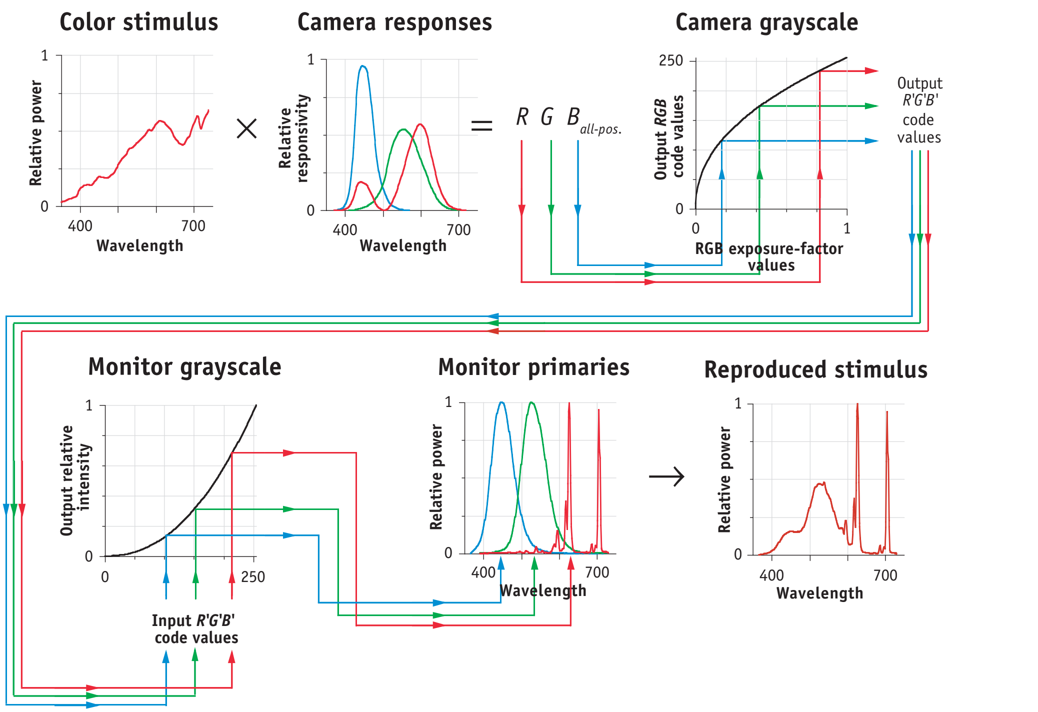


Figure 5.10a Use of a camera having spectral responsivities equivalent to the color-matching functions of the 1931 CIE Standard Colorimetric Observer.

problem would be encountered: each of the monitor color-matching functions is *negative* for certain ranges of wavelengths. (As discussed in Chapter 1, *all* sets of physically realizable primaries have associated color-matching functions with at least some negative regions.) This means that the ideal camera would need to have "negative sensitivity" to light at certain wavelengths!

That is not possible, of course. One could, however, construct a camera having spectral responsivities that correspond to some *other* set of color-matching functions having *no* negative regions. For example, the camera could have spectral responsivities that correspond to the all-positive color-matching functions of the CIE Standard Observer. That essentially would make the camera a colorimeter, which certainly would seem a reasonable thing to do given that the objective is to achieve colorimetrically accurate color reproduction.

Figure 5.10a illustrates the use of such a camera. The figure shows a color stimulus being captured according to the camera's all-positive color-matching-function spectral responsivities, which are normalized such that they are white balanced for the chromaticity of CIE Standard Illuminant D_{65}. The resulting *RGB* exposure-factor values are then processed by the camera to form digital code values. Note that the exposure-factor to code-value relationship of the camera is the same for all three color channels. This is consistent with the fact that D_{65} neutrals produce equal *RGB* exposure-factor values and thus also should produce equal code values. The camera code values are then input to the monitor, which produces corresponding levels of *RGB* relative intensities ("amounts" of the monitor primaries) according to the monitor's code-value to relative-intensity relationship. This relationship also is the same for all three color channels because, as discussed earlier, the monitor was adjusted for constant-chromaticity grayscale tracking. Finally, color stimuli are produced according to the camera *RGB* exposure-factor and code values, the corresponding relative intensities of the monitor primaries, and the spectral characteristics of those primaries.

Again, the objective at this point in the discussion was to achieve colorimetrically accurate color reproduction. However, except for neutral colors, the colorimetry of monitor stimuli produced by this process would *not* match the colorimetry of the corresponding original stimuli. The reason is that an electronic camera having CIE Standard Observer color-matching-function spectral responsivities would generate *RGB* exposure-factor values equal to neutral-normalized CIE *XYZ* values; it would not generate exposure-factor values equal to the monitor *RGB* relative-intensity values required for colorimetric matching (Table 5.3).

The resulting colorimetric errors are shown in Figure 5.10b. In this figure, the CIELAB a^*b^* coordinates for the stimuli of the illuminated test target are compared to those for the corresponding stimuli produced on the monitor. The coordinates for the monitor reproductions generally are closer to the neutral point at the center of the diagram. This means that the reproduced colors will have significantly lower *chromas* than the original colors, i.e., they will look less colorful. This can be seen in the test-target images of Figure 5.11.

There would seem to be a real dilemma here. Colorimetric matching apparently requires an electronic camera to have spectral responsivities proportional to the color-matching functions corresponding to the monitor primaries, but such responsivities are not physically realizable. On the other hand, a camera that has physically realizable spectral responsivities, such as the all-positive responsivities that correspond to the color-matching functions of the CIE Standard Observer, produces large colorimetric errors when used with a real monitor.

There is a solution to this dilemma. Recall from Chapter 1 that all sets of color-matching functions are linear combinations of all other sets. This means that any set can be transformed to another set simply by the use of an appropriate three-by-three matrix

Table 5.3 The *RGB* exposure-factor values, R_{exp}, G_{exp}, B_{exp}, for a digital camera having spectral responsivities equivalent to the color-matching functions of the CIE Standard Observer do not equal the monitor *RGB* relative intensity values, I_R, I_G, I_B, required for colorimetric matching.

	Intensities Required for Colorimetric Matching			Exposures from CIE Standard Observer Responsivites		
Test-Target Color	I_R	I_G	I_B	$R_{all\text{-}pos.}$	$G_{all\text{-}pos.}$	$B_{all\text{-}pos.}$
Neutral 1	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
Neutral 2	0.5890	0.5890	0.5890	0.5890	0.5890	0.5890
Neutral 3	0.3630	0.3630	0.3630	0.3630	0.3630	0.3630
Neutral 4	0.2000	0.2000	0.2000	0.2000	0.2000	0.2000
Neutral 5	0.0890	0.0890	0.0890	0.0890	0.0890	0.0890
Neutral 6	0.0320	0.0320	0.0320	0.0320	0.0320	0.0320
Red	0.3953	0.0156	0.0444	0.2104	0.1180	0.0514
Green	0.0893	0.3209	0.0527	0.1557	0.2380	0.0909
Blue	0.0313	0.0407	0.3322	0.0894	0.0620	0.2819
Cyan	−0.0010	0.2631	0.4038	0.1568	0.2050	0.3710
Magenta	0.4598	0.0748	0.3290	0.3136	0.1970	0.2975
Yellow	0.9037	0.6827	−0.0400	0.6597	0.6820	0.0900

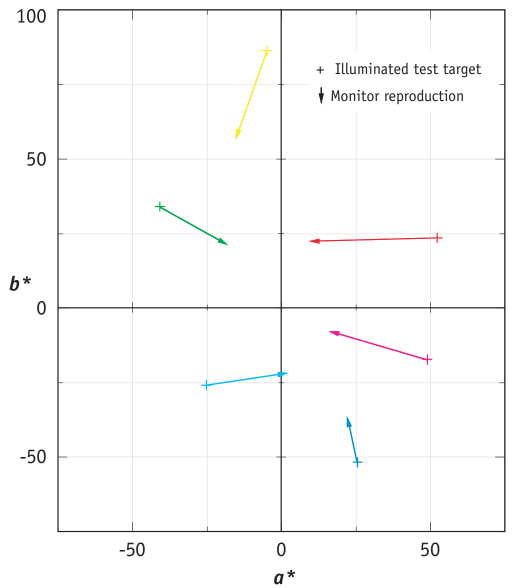


Figure 5.10b The lengths of the vector arrows in this CIELAB a^*, b^* diagram indicate the magnitudes of the chromatic errors that result from using a camera having the spectral responsivities of the CIE 1931 Standard Colorimetric Observer and a monitor having the primaries shown in Figures 5.7a and 5.7b.

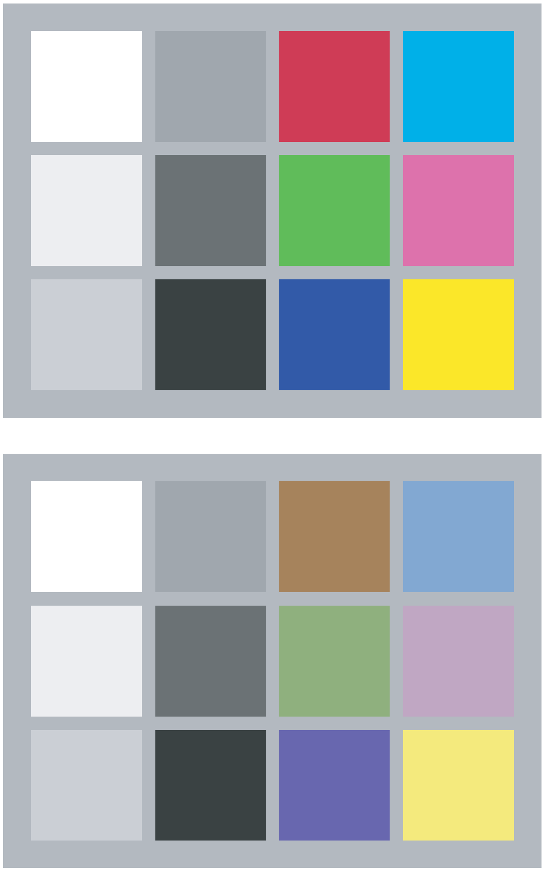


Figure 5.11 The upper image approximates the appearance of the original test colors. The lower image approximates the appearance of a monitor image that results from using a camera having the spectral responsivities of the CIE 1931 Standard Colorimetric Observer and a monitor having the primaries of Figures 5.7a and 5.7b.

multiplication operation. It follows, then, that a matrix can be used to transform the tristimulus values of one set of color-matching functions to those of any other set of color-matching functions.

This means that a real camera *can* have physically realizable spectral responsivities corresponding to the color-matching functions of the Standard Observer or to any of a virtually unlimited number of other sets of all-positive color-matching functions. In all cases, the all-positive color-matching functions will correspond to sets of *imaginary* primaries, but that is perfectly acceptable. A matrix can be used as part of the camera's signal processing to transform its *RGB* exposure signals (i.e., its *RGB* tristimulus values, which correspond to those of the imaginary primaries) to new $R_mG_mB_m$ exposure signals having the appropriate values for the actual monitor primaries (Figure 5.12).

Because device spectral responsivities must be physically realizable, and because the color-matching functions corresponding to actual device primaries always have regions of negativity, it is *always* necessary to create interactions, such as those produced by a matrix, among the color signals of *any* real image-capture device having spectral responsivities equivalent to a set of color-matching functions.

In the example system, where the camera spectral responsivities are equivalent to the all-positive color-matching functions of the CIE Standard Observer, the following matrix transformation can be applied

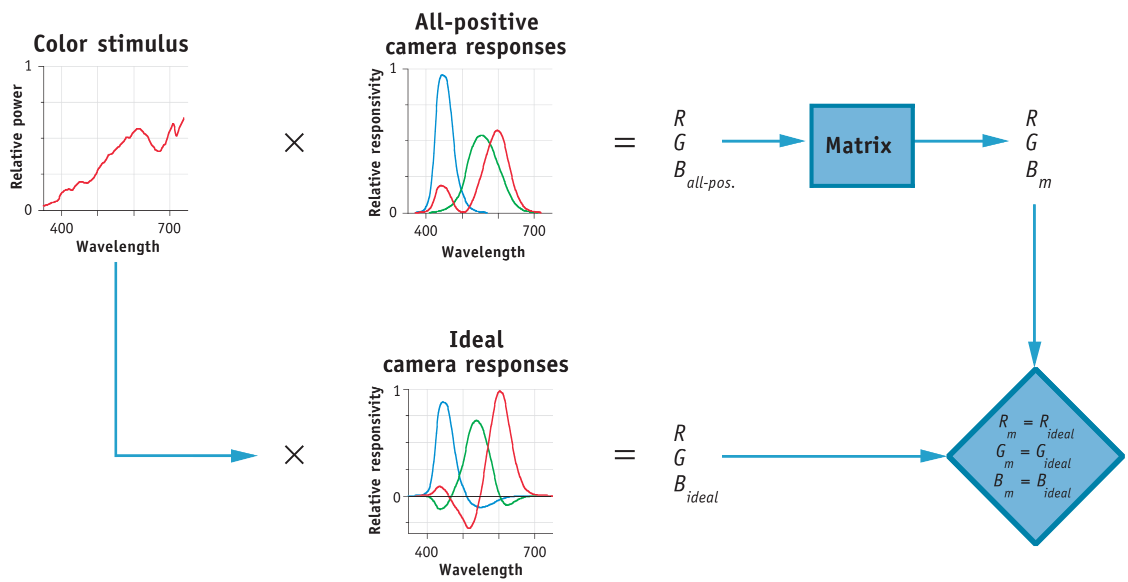


Figure 5.12 Use of a matrix in the signal processing of camera signals. The $R_mG_mB_m$ values of the matrixed signals from the camera having spectral responsivities equivalent to a set of all-positive color-matching functions will exactly equal the $R_{exp}G_{exp}B_{exp}$ values of the signals from the camera having spectral responsivities equivalent to the actual color-matching functions corresponding to the monitor color primaries.

to the D_{65}-balanced *RGB* all-positive camera signals to form new signals, $R_mG_mB_m$:

$$\begin{bmatrix} R_m \\ G_m \\ B_m \end{bmatrix} = \begin{bmatrix} 2.654 & -1.182 & -0.472 \\ -1.078 & 2.040 & 0.038 \\ 0.080 & -0.297 & 1.217 \end{bmatrix} \begin{bmatrix} R_{all\text{-}pos} \\ G_{all\text{-}pos} \\ B_{all\text{-}pos} \end{bmatrix} \tag{5.6}$$

As shown in Table 5.4, the $R_mG_mB_m$ signal values are *identical* to those that would have been produced by a camera having ideal spectral responsivities, i.e., those equivalent to the monitor color-matching functions.

The example system, which now includes a digital still camera having spectral responsivities corresponding to the color-matching functions of the CIE Standard Observer, signal processing that implements the matrix given in Equation (5.6), and nonlinear signal processing that produces a one-to-one grayscale reproduction on a monitor having the primaries of Figures 5.7a and 5.7b, produces the results given in Figure 5.13. These results show that, for the most part, the colorimetry of the illuminated test target has been reproduced exactly on the monitor. The colorimetry of the yellow patch, however, has not been reproduced accurately. That is because the monitor physically cannot generate a stimulus having that particular colorimetry; that stimulus is outside the monitor's *color gamut*. This can be seen in the required intensity values, I_R, I_G, *and* I_B, given in each of the tables. The required blue intensity value for the yellow color is negative. To display this color, then, the monitor would have to produce *negative* amounts of blue light! Obviously it cannot do that, so the monitor cannot produce a colorimetric match to that particular yellow color.

Results of colorimetric matching

So far, the example camera/monitor combination produces a *perfect* colorimetric match of any color captured by the camera, as long as that color is within the monitor's color gamut. It would seem, then, that this system should produce excellent images. The system does, in fact, make very good monitor

Table 5.4 The matrixed exposure-factor values, R_m, G_m, B_m, for a digital camera having spectral responsivities equivalent to the color-matching functions of the CIE Standard Observer exactly equal the monitor *RGB* relative intensity values, I_R, I_G, I_B, required for colorimetric matching. Use of the matrix produces exposure-factor values that are identical to those that would be produced by a camera having ideal spectral responsivities.

	Intensities Required for Colorimetric Matching			Matrixed Exposures from CIE Standard Observer Responsivites		
Test-Target Color	I_R	I_G	I_B	R_m	G_m	B_m
Neutral 1	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
Neutral 2	0.5890	0.5890	0.5890	0.5890	0.5890	0.5890
Neutral 3	0.3630	0.3630	0.3630	0.3630	0.3630	0.3630
Neutral 4	0.2000	0.2000	0.2000	0.2000	0.2000	0.2000
Neutral 5	0.0890	0.0890	0.0890	0.0890	0.0890	0.0890
Neutral 6	0.0320	0.0320	0.0320	0.0320	0.0320	0.0320
Red	0.3953	0.0156	0.0444	0.3953	0.0156	0.0444
Green	0.0893	0.3209	0.0527	0.0893	0.3209	0.0527
Blue	0.0313	0.0407	0.3322	0.0313	0.0407	0.3322
Cyan	−0.0010	0.2631	0.4038	−0.0010	0.2631	0.4038
Magenta	0.4598	0.0748	0.3290	0.4598	0.0748	0.3290
Yellow	0.9037	0.6827	−0.0400	0.9037	0.6827	−0.0400

images *of the test target*. It also makes good images when the camera is aimed at a *reproduction*, such as a reflection print. However, monitor images of *live* scenes, especially outdoor scenes, captured by the camera look quite “flat” (Figure 5.14). Specifically, they appear too low in luminance contrast and too low in color saturation.

What is wrong? The monitor images, while colorimetrically accurate as measured by the colorimeter, are poor in color quality because several important factors have not yet been accounted for in the system. One of these factors is physical; the others are perceptual.

Viewing flare

The physical factor not yet accounted for in the example system is *viewing flare*. When an image is viewed on a monitor, ideally the observer should see only the light emitted by the monitor itself. But in most practical viewing situations, the observer also will be seeing flare light (stray light) that is reflected from the monitor’s glass faceplate. Flare light can come from ordinary light sources in the viewing environment (overhead lamps, window light, etc.). It also can come from the monitor itself—if, for example, the observer is wearing a white shirt that reflects monitor light back to the faceplate.

Flare light is added, more or less uniformly, to the light from the image being displayed on the monitor. Therefore, the luminance at each point in the system grayscale will be increased additively. This is shown, in terms of linear luminance units, in Figure 5.15a. The representation in this figure might seem to imply that the effect of the flare light should be virtually undetectable. In reality, however, the effect will be very noticeable.

The visual impact of the effect is much better represented when the system grayscale is expressed in terms of negative log luminance-factor values (Figure 5.15b). Again, such values are more consistent with the way the human visual system responds to variations in luminances. As the figure shows, the addition of flare light significantly brightens darker areas of an image, but it has a much less apparent effect on brighter areas. Therefore, the luminance

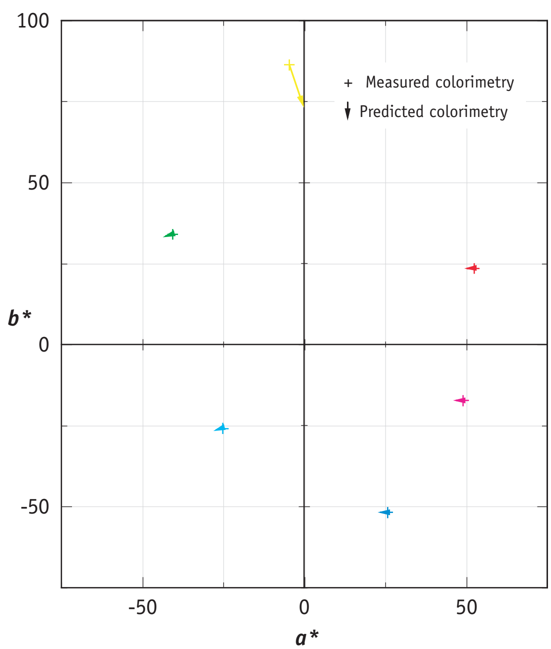


Figure 5.13 Most colorimetric errors shown in Figure 5.10b can be eliminated by the inclusion of an appropriate signal processing matrix. The yellow color is outside the reproducible color gamut of the example monitor.

contrast of a displayed image is reduced, particularly in the darker areas.

Because flare light generally is white or nearly so, its addition also will reduce the saturation of colors, as shown in Figures 5.16a and 5.16b. In Figure 5.16a, the test colors have been photographed as if they were part of the *principal subject area* of the original scene, i.e., as if they were in the area used in determining the overall camera exposure. The test colors therefore are normally exposed. In Figure 5.16b, the test colors have been photographed as if they were in a deeply shaded area of the same original scene. When photographed that way, the test colors are underexposed and effectively become darker. As the figures show, the desaturation caused by viewing flare is much greater for the darker test colors. (Refer to Appendix E for more details on viewing flare.)

In order to determine the proper compensation for the effects produced by viewing flare, the amount of flare light in the monitor viewing environment first must be measured. There are various methods that can be used. In one, illustrated in Figure 5.17, an opaque white patch having reflective properties comparable to those of the monitor's faceplate is attached to the faceplate, and the amount of light reflected from that patch is measured from the observer's viewing position. Since the patch is opaque, any light measured from it must be flare light. If no image is present on the monitor, the measurement will represent flare due to other sources of light in the environment. If an image is present, light from the monitor itself will contribute to the total flare as that light reflects from room objects (including any observers) and falls back to the monitor's faceplate. Flare-light measurements should be made using an appropriate instrument, such as a telephotometer, designed to measure a small area of light from a distance. The measured amount of flare light usually is expressed as a percentage of the light measured, using the same type of instrument, from a white area of a representative image displayed on the monitor.

In viewing environments where some care is taken to eliminate obvious sources of stray light, the luminance of the flare light generally will be equal to about 0.5 to 1.0 % that of a displayed image white. Although that might sound like a small amount, the effect of that much flare light is *very* significant. It *must* be compensated for if high-quality images are to be produced.

Signal processing can be used to compensate, at least partially, for the effects of flare light. The compensation, shown in Figure 5.18, produces an exponential increase in the negative log luminances of the darker regions of the system grayscale. That should be expected. Flare is an addition of light, which requires a subtraction of light for compensation. When plotted on a logarithmic scale, the linear subtraction will produce the exponential logarithmic increase shown in the figure.

With appropriate signal processing, *RGB* image signals can be adjusted such that colorimetric measurements made *from the observer's position* (as shown in Figure 5.17) have a one-to-one relationship to measured luminance-factor values of the original scene. This result can be achieved in practice except for the darkest colors. When that signal processing is applied, the practical system grayscale *measured in the absence of flare light* will be similar to that shown in Figure 5.18.

Figure 5.14 The upper image approximates the appearance of a monitor image that accurately matches the colorimetry of the original scene. The lower image approximates the appearance of a monitor image that is not colorimetrically accurate, but which generally would be judged to have improved color reproduction.

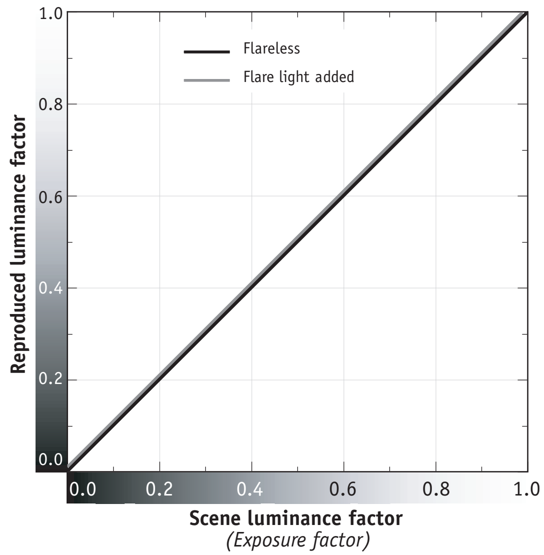


Figure 5.15a Effect of flare light on the system grayscale characteristic, expressed in terms of linear reproduced luminance-factor values.

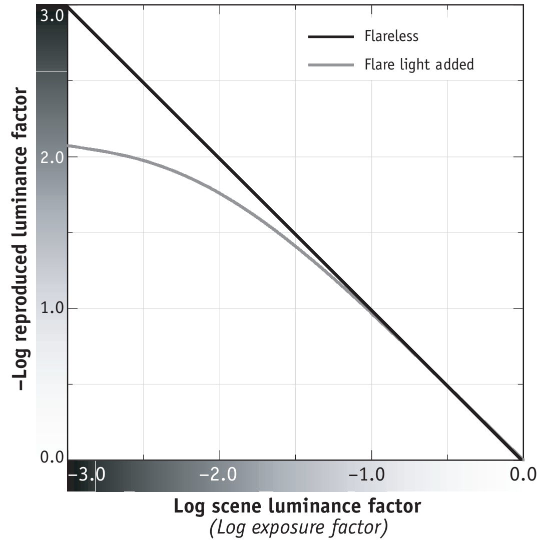


Figure 5.15b Effect of flare light on the system grayscale characteristic, expressed in terms of negative log reproduced luminance-factor values. The visual impact of this effect is more appropriately represented by these logarithmic values. Note that the effect is greatest for blacks and dark grays.

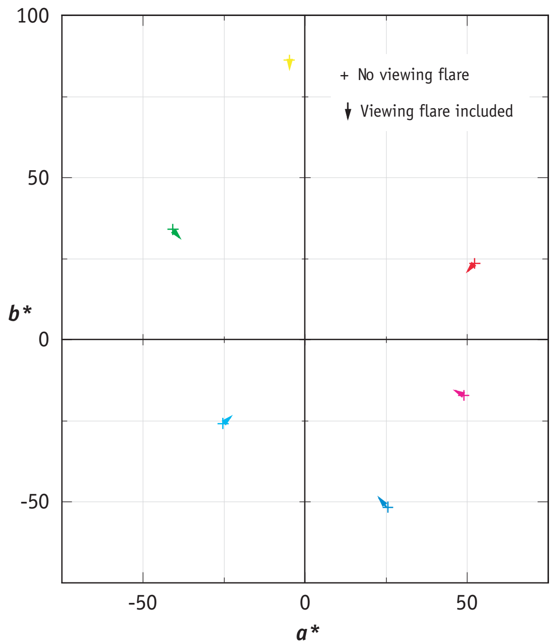


Figure 5.16a Effect of flare light on color saturation. Since flare light essentially is white, its addition reduces the saturation of colors, i.e., it moves them toward white. In this figure, the test colors have been photographed as part of the principal subject area of an original scene.

Perceptual factors

With the addition of flare-compensating signal processing in the example system, monitor images from live scenes are improved greatly. However, to most observers, the monitor images still appear somewhat low in luminance contrast and color saturation. This result might seem surprising, because measurements made from the observer's position confirm that colorimetric matching has indeed been achieved. Nevertheless, as stated earlier, when such discrepancies arise, the judgments of the observers must prevail. So despite the fact that colorimetric matching has been achieved in the system, the remaining problems with the appearance of images must be addressed. If these problems are not due to physics, they must be due to perceptual phenomena.

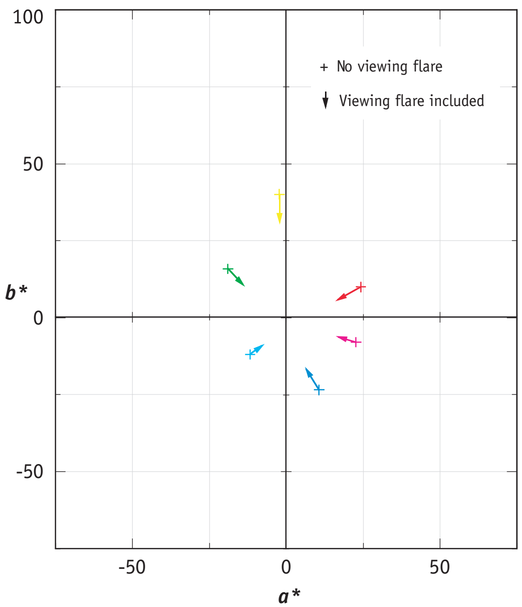


Figure 5.16b Effect of flare light on the color saturation of very dark colors. In this figure, the test colors have been photographed as if they were in a deeply shaded area of the original scene. The desaturation caused by viewing flare is much greater for these darker test colors.

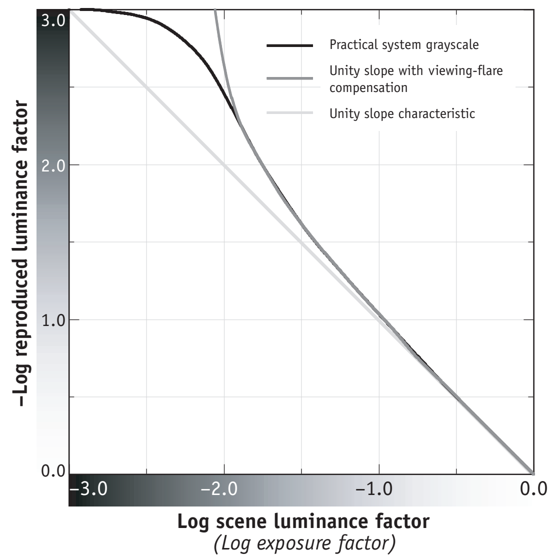


Figure 5.18 A practical system grayscale characteristic designed to compensate partially for the addition of 0.75 % viewing-flare light. The reproduced luminance-factor values are measured in the absence of flare light. Also shown for reference are a straight line of unity slope and a theoretical compensation for 0.75 % viewing flare.

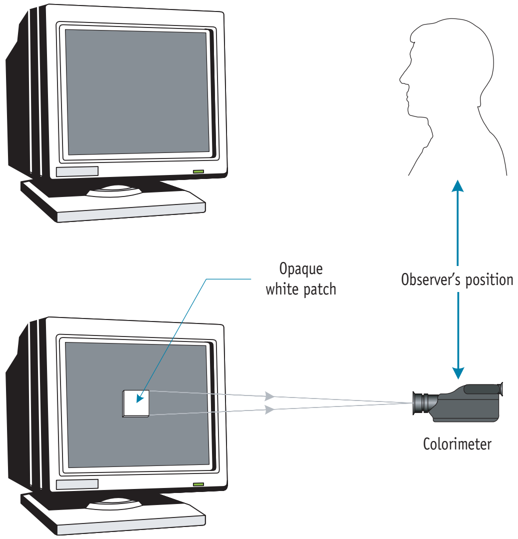


Figure 5.17 One method for measuring flare light reflected from the face of a CRT.

There are several possible factors that might contribute to the apparent discrepancy between color measurements made of the images and their color appearance. A monitor image is somewhat less sharp than the original, and it is known that a reduction in image sharpness can produce a corresponding reduction in perceived image luminance contrast. A reduction in image sharpness also can produce a reduction in perceived image color saturation.

In addition, the *absolute* luminances produced by the monitor generally will be significantly lower than those of the original scene. For example, the absolute luminances of an outdoor scene may be hundreds of times greater than those that can be produced by a reproduction of that scene on a typical high-resolution computer monitor. As discussed in Chapter 3, images having lower absolute luminances are perceived to be lower in luminance contrast, and they also are perceived to be lower in colorfulness.

Another factor is that, in most cases, the monitor is not viewed simultaneously with the original scene. The monitor image therefore will be judged according to the observer's *memory* of the original, and colors generally are remembered as being somewhat more saturated than they really were. Finally, many observers *prefer* colors to be reproduced even *more* saturated than they remember them being. For example, the preferred reproduction of the color of the sky near the horizon closely corresponds to the much more saturated color of the actual sky directly overhead.

There are a number of factors, then, to suggest that the color signals of an electronic camera should be modified somewhat in a way that will increase both the luminance contrast and color saturation of images produced on the monitor. But how should these modifications be made, and how can their magnitudes be determined?

Broadcast televisions signals

There are several techniques that could be used to determine, from first principles, the signal processing required to compensate for the perceptual factors described above. Another approach would be to examine, or "reverse engineer," a successful existing system to see how such modifications are performed in actual practice. That is the approach that will be taken here.

A television broadcast system is a reasonable choice for examination, since broadcast specifications and practices are based on decades of industry experience. Our own experience also supports this choice. We have found that although some improvements certainly can be made, the color quality of electronically displayed images generally is quite high when the relationship between original and reproduced colorimetry is essentially that produced by conventional broadcast television systems.

Figure 5.19 shows the measured grayscale characteristic for a representative broadcast system. Both the camera and the monitor have been adjusted according to broadcast industry recommendations. The grayscale characteristic results from the use of a monitor similar to that of the example system and a broadcast television camera employing signal processing corresponding to the following equations:

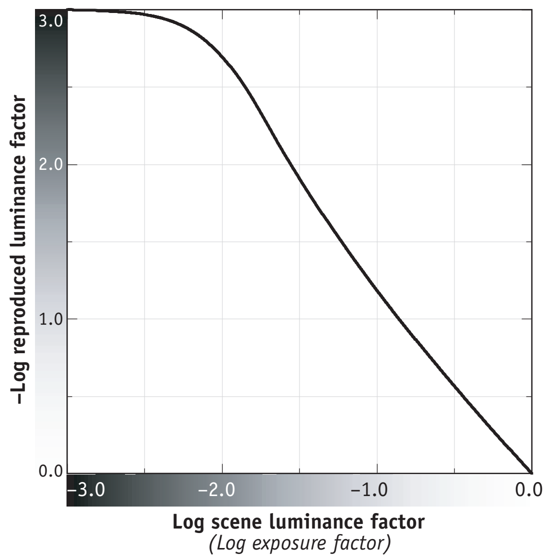


Figure 5.19 Measured system grayscale characteristic for a broadcast television system.

for $0.018 \le Y \le 1.0$:

$$V = 1.099Y^{0.45} - 0.099 \tag{5.7a}$$

for $0.0 \le Y < 0.018$:

$$V = 4.5Y \tag{5.7b}$$

where V is the camera output signal or voltage, and Y is the scene luminance-factor value.

These equations, which are from ITU-R BT.709, *The HDTV Standard for the Studio and for International Programme Exchange*, originally were derived from analog measurements made of broadcast television cameras that had been adjusted by skilled camera operators such that excellent quality images were produced on studio CRT monitors set up and viewed according to broadcast recommendations.

The grayscale characteristic of an electronic camera having signal processing corresponding to these equations is *not* simply a mathematical inverse of the CRT monitor characteristic. This is shown in Figures 5.20a and 5.20b, which compare the grayscale characteristics for two digital cameras. The signal processing of one camera is an exact inverse of the monitor characteristic, while that of the other camera

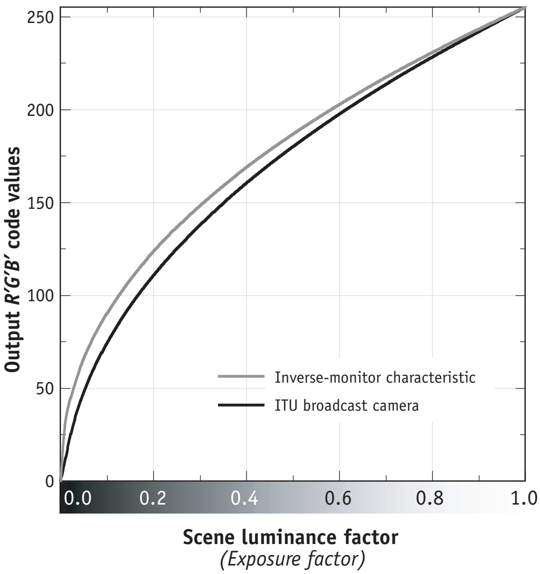


Figure 5.20a Digital camera characteristics. The signal processing of one camera is a mathematical inverse of the monitor characteristic. The signal processing of the other corresponds to the ITU broadcast camera equations.

corresponds to the ITU equations. The camera characteristics are normalized for white and expressed in terms of output code values as a function of linear scene luminance-factor value (Figure 5.20a) and log scene luminance-factor value (Figure 5.20b).

The overall system grayscales produced by these cameras, when used with the CRT monitor of Figure 5.6, are shown in Figure 5.21. As would be expected, the camera having the monitor-inverse characteristic produces a straight-line overall system grayscale of unity slope. The camera corresponding to the ITU equations produces a system grayscale consistent with broadcast television.

In the example system, the monitor images are significantly improved by the use of camera signal processing corresponding to the ITU equations. This improvement is illustrated in the images of Figure 5.22. There are several factors that contribute to this improvement.

First, the use of signal processing corresponding to the ITU equations produces a system grayscale characteristic that compensates partially for about 0.75 % viewing flare (Figure 5.23a). That amount

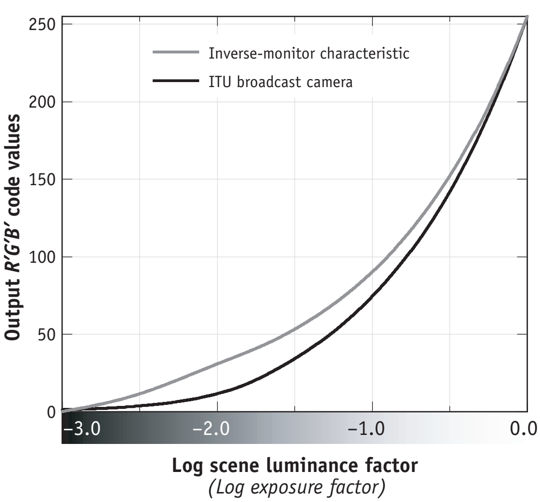


Figure 5.20b Characteristics for the cameras of Figure 5.20a, expressed in terms of output $R'G'B'$ code values as a function of log scene luminance-factor value.

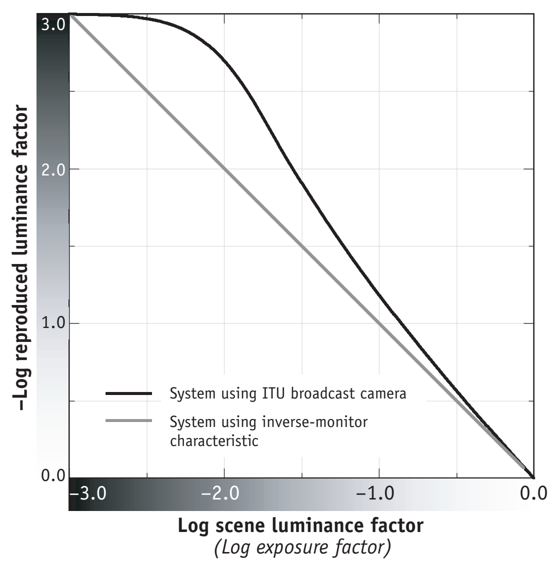


Figure 5.21 Overall system grayscale characteristics resulting from the use of the cameras of Figure 5.20a (and Figure 5.20b), and the monitor of the example system.

Figure 5.22 Approximate appearances of monitor images corresponding to the overall system grayscale characteristics shown in Figure 5.21. The upper image has a straight-line system grayscale of unity slope. The lower image has a system grayscale consistent with broadcast television practices. That broadcast system grayscale also was used to produce the lower image of Figure 5.14.

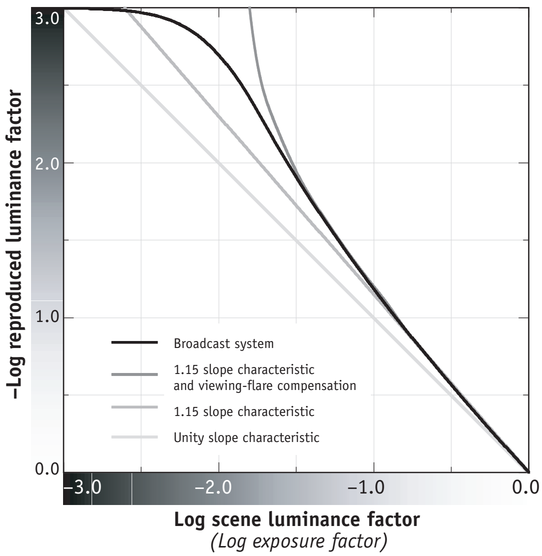


Figure 5.23a A system grayscale characteristic resulting from ITU camera signal processing. Also shown are a unity slope characteristic, a 1.15 slope characteristic, and a 1.15 slope characteristic with 0.75 % viewing-flare compensation.

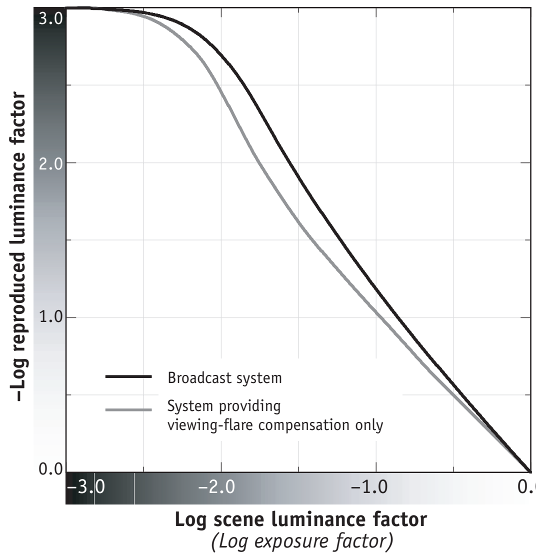


Figure 5.23b Comparison of system grayscale characteristics.

of flare corresponds well with the example conditions, and it is typical for many electronic display viewing environments. The figure also shows that, in addition, the camera signal processing increases the overall slope of the system grayscale by about 15 %.

It was this slope increase that was lacking in the system grayscale characteristic examined earlier (Figure 5.18). That grayscale provided flare compensation only. The two system grayscale characteristics are compared in Figure 5.23b. The slope increase in the broadcast system is desirable because it helps compensate for the reduction in perceived luminance contrast resulting from the relatively low absolute luminance levels of the monitor.

Because the camera signal processing is applied to camera *RGB* signals (not just to a separate achromatic signal), the slope increase and flare compensation also affect overall color reproduction. The reason for this, as shown in Figure 5.24, is that in an additive system, a grayscale is formed by the addition of separate red, green, and blue color scales. The signal processing used to alter the grayscale does so by altering these individual color scales, and the higher color contrasts produced by these alterations result in higher overall reproduced color saturation (Figure 5.25). That is desirable, because higher saturation is needed in order to compensate for the color saturation decreases associated with viewing flare and with the various perceptual factors described earlier.

Some textbooks and other publications suggest that a slope greater than unity—often called *system gamma*—of a television system grayscale characteristic is required because television displays generally are viewed in a dim-surround viewing environment, i.e., an environment in which the displayed image is somewhat brighter than the areas immediately surrounding the display. However, our experience is that when the original is a live scene, red, green, and blue slope increases of 15 to 20 % (in addition to any slope increases required for viewing-flare compensation) are desirable on all electronic display systems, including systems where the display is viewed in *average-surround* conditions. This is particularly true when the original subject is a highly illuminated outdoor scene.

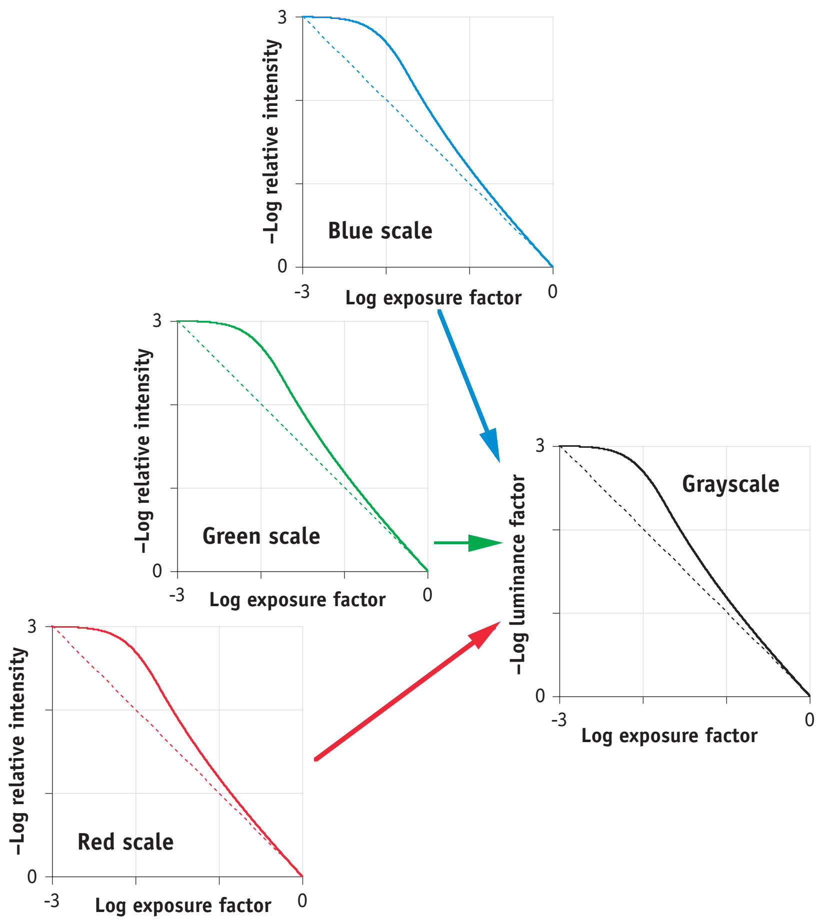


Figure 5.24 A monitor grayscale composed of red, green, and blue light. If the grayscale shape is altered (e.g., from the dotted unity line to the solid curve) using RGB signals, the individual red, green, and blue color scales also are affected.

Furthermore, if these desirable slope increases were related only to a dim-surround effect, then only luminance values would need to be altered. (The presence of a dim surround has been shown to affect perceived image luminance contrast almost exclusively.) However, our experience is that higher-quality images invariably are produced when the increase is applied to each of the *RGB* signals, rather than to just a luminance signal. Increasing the slope of the *RGB* signals results in increases in luminance contrast *and* color saturation, which suggests that the adjustment is required for reasons other than dim-surround correction. As discussed earlier, the possible reasons include compensation for the relatively low absolute luminances of the monitor, lack of image sharpness, color-memory effects, and color-reproduction preferences.

What is true, however, is that in some viewing conditions, including most home television viewing, an *additional* increase in luminance contrast (but not color contrast) is required in order to compensate for dim-surround viewing. That additional luminance contrast requirement is consistent with the higher-slope grayscale characteristics, compared

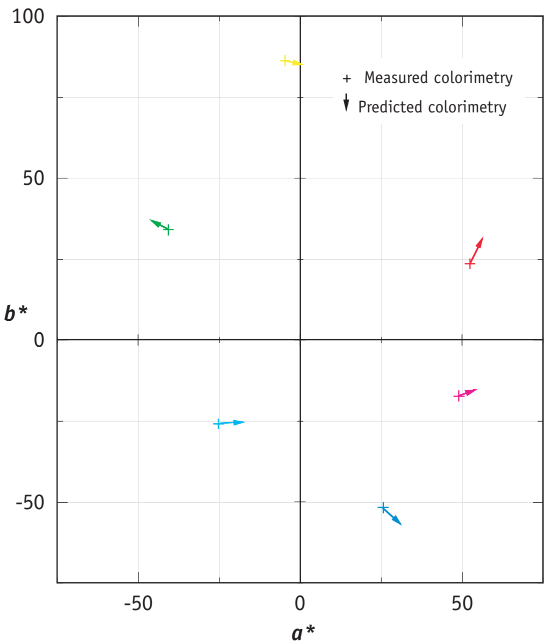


Figure 5.25 Chroma increases associated with the 15 % slope increase and viewing-flare compensation resulting from the use of ITU signal processing. The a^*b^* values are based on measurements made in the absence of flare light.

to those of broadcast studio and computer monitors, often measured on home television displays. In television systems, this increase can be made conveniently at the display, where the signal processing is arranged such that luminance contrast can be increased without increasing the chromas of colors.

An ideal electronic imaging system

From this analysis, it can be seen that in an ideal electronic imaging system (Figure 5.26), the camera's image-capture characteristic would consist of spectral responsivities equivalent to an all-positive set of color-matching functions. The camera signal processing would include a matrix to transform its linear image-capture color-signal values to those that would have been formed if the camera had spectral responsivities equivalent to the particular color-matching functions for the monitor primaries. Nonlinear signal processing also would be used in the ideal camera to produce color signals that, when used by the monitor, would produce desirable colorimetric departures from a one-to-one colorimetric reproduction of the original scene.

Although the responsivities shown in Figure 5.26 are equivalent to the color-matching functions of the CIE 1931 Standard Colorimetric Observer, it generally is more practical to use responsivities equivalent to other sets of all-positive color-matching functions. Appropriately chosen responsivities, such as those shown in Figure 5.27, can reduce the magnitudes of the signal processing matrix coefficients. This is important, because most forms of electronic noise increase significantly as the magnitudes of the matrix coefficients increase.

Actual digital still cameras and other types of electronic cameras conform to this ideal to varying degrees. While perhaps no real camera has spectral responsivities that are exactly equivalent to a set of color-matching functions, some come reasonably close. Most electronic cameras incorporate matrices to transform their linear *RGB* signals based on reference color primaries defined by industry standards. Virtually all electronic cameras produce nonlinear output signals, again in accordance with industry standards, which, when used with an appropriately set up color monitor, produce an overall system grayscale characteristic closely approximating that previously shown in Figure 5.21. Some cameras also produce "raw" image files containing minimally processed data from the camera's image sensor. Raw files generally are proprietary and may differ among manufacturers and even among camera models from the same manufacturer. If sufficient information is available regarding such files, subsequent signal processing can be used to apply the matrix and nonlinear transformations required for proper display.

Color-encoding considerations

This analysis has produced two important outcomes that directly relate to color encoding. First, it has shown that the optimum reproduction of a live original scene is *not* a one-to-one reproduction of the

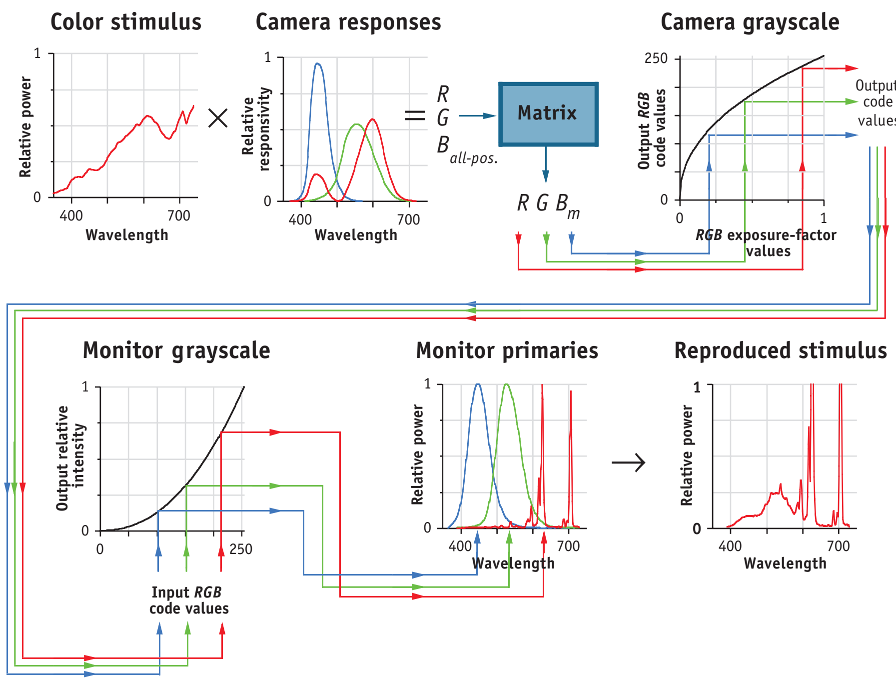


Figure 5.26 An ideal electronic system. The system includes a camera having spectral responsivities equivalent to a set of all-positive color-matching functions. It also includes appropriate signal processing of the camera *RGB* exposures.

scene colorimetry. That result is not unique to electronic color-imaging systems. It is just one illustration of the fact that in order to produce a visual match under most circumstances, the colorimetry of the reproduction must differ from that of the original.

The analysis also has shown that image-processed electronic camera signal values themselves do not correspond directly to original-scene colorimetric values, nor do they correspond directly to reproduced colorimetric values. However, image-processed camera signal values can be color encoded in terms of colorimetric values if their relationships to original-scene colorimetry or to colorimetry reproduced on an electronic display are determined. This can be done in a number of ways. For example, the relationship between original-scene colorimetry and a particular camera's output code values can be defined from knowledge of that camera's spectral responsivities and its image signal processing. The relationship also can be determined empirically using appropriate characterization procedures. Such procedures generally include the use of color test targets and measurement methods similar to those described in this chapter.

The relationship between an electronic camera's $R'G'B'$ signal values and the colorimetry that is produced on an electronic display also can be defined using appropriate characterization procedures. For example, arrays of camera output code values can be input to a display, and the colorimetry of the resulting stimuli can then be measured directly. Alternatively, the relationship can be determined based on knowledge of the output signal processing and the chromaticities of the display primaries. A similar relationship between electronic camera signal values and output colorimetry can be derived based on the properties of a mathematically defined reference display.

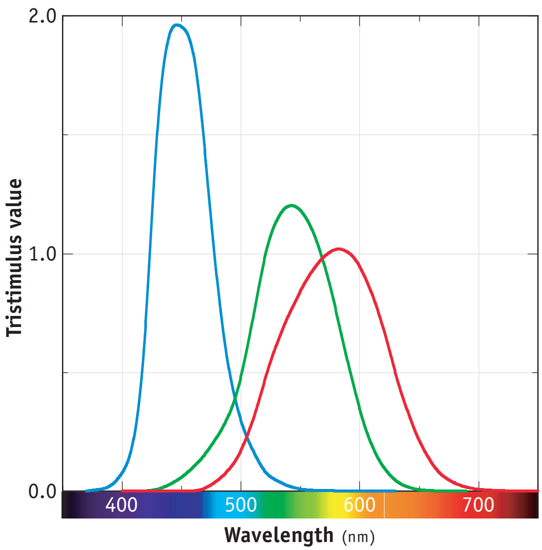


Figure 5.27 These spectral responsivities are equivalent to a set of color-matching functions different from those that define the CIE 1931 Standard Colorimetric Observer. Use of spectral responsivities such as these can reduce the magnitudes of the signal processing matrix coefficients.

The techniques that have just been described make it possible to color-encode electronic camera signals either in terms of the colorimetry of the original scene captured by the camera or in terms of the colorimetry that would be produced if those signals were sent to an actual or reference display device. Another alternative, of course, is simply to encode camera signals directly in terms of their (nonlinear) $R'G'B'$ signal values, without transformations to colorimetric values. Similarly, if linear RGB signal values are provided by the camera, the encoding can be directly in terms of those values. Actual linear or nonlinear camera signal values also can be transformed to those that would have been formed by a specified reference camera, and the reference camera signal values can be used in the encoding process. Each of these types of color encoding can be valid, depending on the requirements of the particular application.

The decision as to which type of color encoding should be used for a given color-imaging system is not always a simple one to make, as will be discussed extensively in later parts. Certainly the particular characteristics of any electronic cameras and electronic displays that are to be supported by a system must be considered. Another important factor that would influence the choice of color encoding would be the colorimetric characteristics of any hardcopy input or output media that a system also might be required to support. The colorimetric characteristics of representative hardcopy media and other factors that must be considered in their color encoding are discussed in the next three chapters of this part.

Summary of key issues

- An ideal image-capture device would have spectral responsivities equivalent to the color-matching functions of the system display primaries.
- Because the color-matching functions of all real output devices have negative values at some wavelengths, the equivalent spectral responsivities are not physically realizable.
- Image signals equal to those that would be produced from physically unrealizable spectral responsivities can be achieved by applying an appropriate matrix transformation to signals produced by an actual image capture device having spectral responsivities equivalent to a set of all-positive color-matching functions.
- Matrixed signals will have positive values for colors that are within the color gamut of the associated display device. Colors outside this gamut will produce negative values in at least one of the color signals. Negative signal values generally will be clipped at a typical display device.
- For historical reasons, the color matrixing and other color signal processing transformations used in electronic cameras often are based on the properties of CRT display devices.
- The optimum displayed reproduction of a live original scene is not a one-to-one reproduction of the scene's actual colorimetry.
- An electronic imaging system's grayscale characteristic must depart from a one-to-one colorimetric

relationship with the original scene in order to compensate for viewing flare and to account for the effects resulting from various perceptual factors.

- Signal-processed electronic camera signals themselves generally do not correspond directly to original-scene colorimetric values or to reproduced colorimetric values.
- For color-encoding purposes, the relationships among scene colorimetry, electronic camera signals, and displayed colorimetry must be defined.

6

Reflection Images

Scans of hardcopy images—such as reflection prints, photographic slides, still and motion picture photographic negatives, and motion picture prints—are an important input source of color information in many color-imaging systems. This chapter begins an analysis of the nature of such images with an examination of the colorimetric characteristics of images produced on reflection supports.

Reflection images can be produced by conventional and instant photography, graphic arts printing, thermal dye transfer, inkjet printing, electrophotography, and by a number of other technologies. However, regardless of the exact technology employed in their production, virtually all pictorial reflection images share certain fundamental characteristics and colorimetric relationships to original scenes. Those characteristics and relationships will be examined in the context of an experimental color-imaging system (Figure 6.1) consisting of a digital camera, a computer, a thermal dye-transfer printer, and a scan printer that writes directly onto conventional photographic paper. The system produces very high-quality images on both the thermal and photographic reflection media.

The test target that was described in Chapter 5 will again be used in the analysis of the system. As before, the target will be illuminated by a light source simulating CIE Standard Illuminant D_{65}. The test-target colorimetry will be computed from the measured spectral reflectances of its patches and the spectral power distribution of the light source. Images reproduced on the reflection media will be illuminated by a different light source—one simulating CIE Standard Illuminant D_{50} (Figure 6.2) which is an industry standard for indoor viewing of reflection images. The situation is both realistic and common; it corresponds, for example, to photographing scenes outdoors and subsequently viewing the reproduction indoors.

The illuminated reproductions will be measured using a colorimeter positioned such that the measurements are free of flare light. The measured colorimetric values of the illuminated reproductions then will be compared to those computed for the illuminated test target.

Reflection media: general characteristics

Figure 6.3 illustrates the basic light-modifying properties of a reflection medium consisting of a colorant (dye, ink, or pigment) and a reflective support. Light incident on the medium is absorbed at least twice by the colorant: once as that light passes through the colorant to the support and a second time after it reflects from the support and passes through the colorant again. Light reflecting from the support generally will be scattered in all directions, so some of that light will be subject to various types of internal reflections before leaving the medium.

Digital Color Management: Encoding Solutions, Second Edition Edward J. Giorgianni and Thomas E. Madden
 ISBN: 978-0-470-51244-9

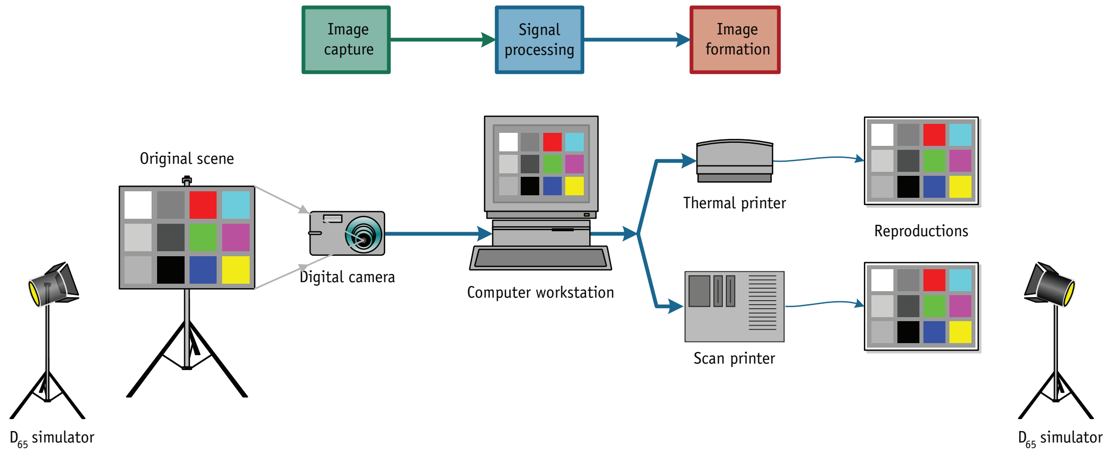


Figure 6.1 Image capture, signal processing, and image formation in an experimental reflection-print color-imaging system.

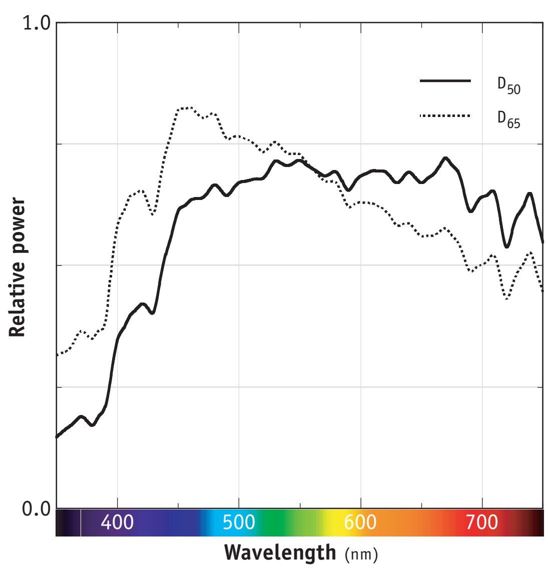


Figure 6.2 Relative spectral power distributions for CIE Standard Illuminants D_{50} and D_{65}.

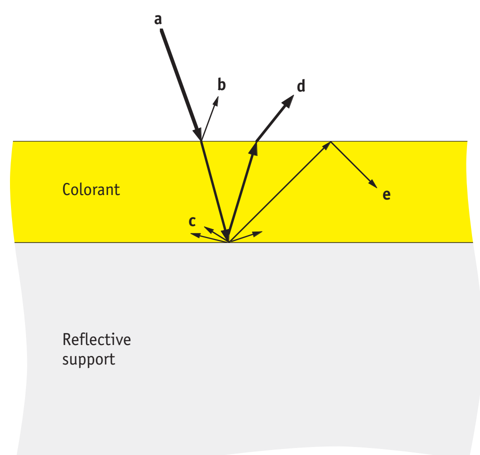


Figure 6.3 Reflection optics. A fraction of the light (a) incident on a reflection medium may reflect from the front surface (b). Most incident light will pass through the colorant to the support (c), where it will be scattered in all directions. Light reflected from the support will again pass through the colorant to the front surface, where it may exit (d) or be internally reflected (e). Depending on the nature of the medium, other types of internal reflections also may occur.

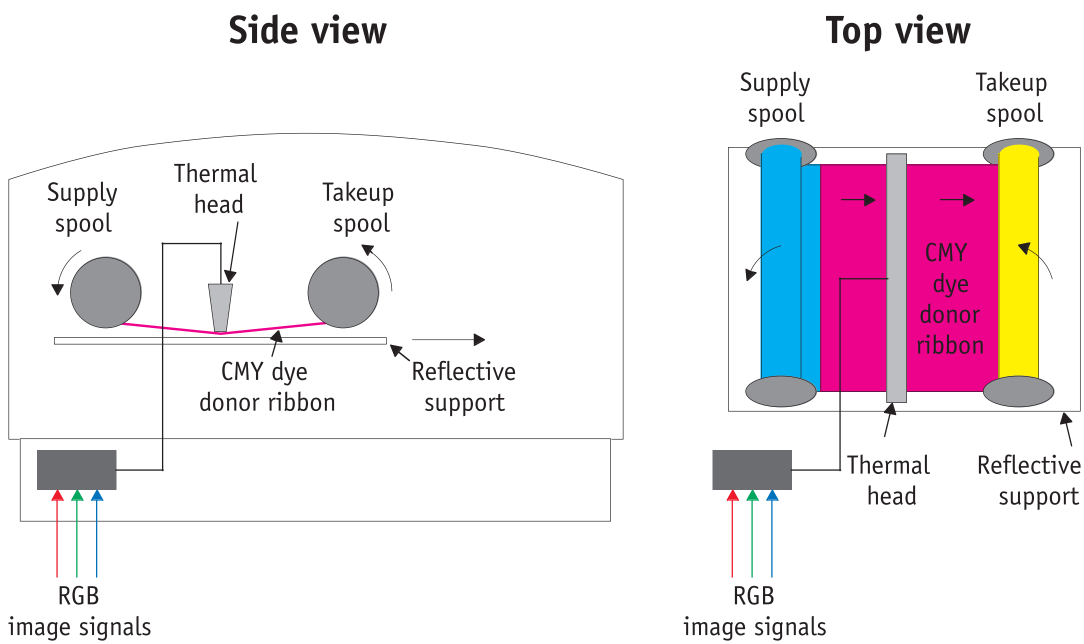


Figure 6.4a In a thermal printer, cyan, magenta, and yellow dyes are transferred sequentially from a donor to a reflective support.

Some incident light may be reflected from the front surface of the medium without passing through the colorant. Such front-surface reflections cause flare in reflection-image viewing, and that flare must be accounted for in the design of the imaging system.

Reflection color-imaging media use multiple colorants to selectively absorb particular wavelengths of light. In the three-colorant media that will be discussed here, a cyan dye is used to absorb red light, a magenta dye to absorb green light, and a yellow dye to absorb blue light. In a thermal printer, these image-forming dyes are transferred sequentially from a donor to a reflective support (Figure 6.4a). The amount of each dye that is transferred is determined by the red, green, and blue image signals sent to the printer.

Conventional photographic papers also use cyan, magenta, and yellow image-forming dyes to absorb red light, green light, and blue light, respectively. These dyes are formed within three separate layers that are coated on a reflective support (Figure 6.4b). The amount of each dye that is formed is determined by the amount of exposure received in each layer. The exposing light may come from an optical device, such as an enlarger used in printing photographic negatives. In the experimental system described here, the exposure is provided by a scan printer. The amount

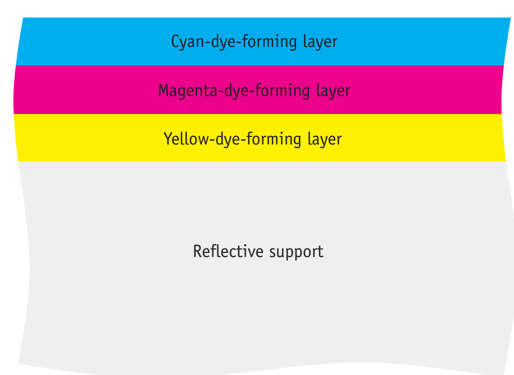


Figure 6.4b In photographic paper, cyan, magenta, and yellow dyes are formed within light-sensitive layers coated on a reflective support.

of each dye that is formed is controlled by the red, green, and blue image signals sent to the printer.

Reflection neutrals: colorimetric considerations

In the electronic camera and display system examined in Chapter 5, neutrals were defined as stimuli having chromaticity coordinates identical to those of the original-scene illuminant. That is why the display white point was set to those chromaticities. However, the experimental reflection-print system being studied here is different in two respects.

First, unlike a self-luminous display, a reflection image alone cannot generate color stimuli. A reflection image is an *object* that forms stimuli only when illuminated. A viewing illuminant therefore is required, and the spectral power characteristics of the resulting stimuli will be influenced directly by the spectral power characteristics of that illuminant.

Second, there are *two* illuminants used in the experimental system: a D_{65} source for illuminating the original scene, and a D_{50} source for illuminating the reflection image. The chromaticities of the two illuminants are different ($x = 0.3127$, $y = 0.3290$ for D_{65}; $x = 0.3457$, $y = 0.3585$ for D_{50}). This chromaticity difference, and the effect that difference will have on the observer, must be taken into account when evaluating neutrals reproduced by the system.

For example, suppose the spectral reflectances of the system's reproduced neutrals were such that when they were illuminated by D_{50} their resulting luminance factors and chromaticity coordinates were equal to those of the original D_{65}-illuminated test-target neutrals (Figure 6.5). Although such reproduced neutrals would be exact *colorimetric* reproductions of the target neutrals, they would not *appear* achromatic. They would appear too blue.

The reason for this is that an observer's state of chromatic adaptation will be different in the original-scene and reproduced-image viewing environments. As described in Chapter 3, chromatic adaptation is an adjustment of the visual mechanism in response to the average chromaticity of the stimuli being viewed. An area of a reflection image ordinarily must have the same chromaticity *as that of the viewing illuminant itself* in order to appear achromatic. To appear achromatic when viewed in the D_{50}-illuminated viewing environment, then, the reproduced neutrals must have the chromaticity of D_{50}, not of D_{65} (Figure 6.6).

In the experimental system, reflection reproductions of the test-target neutral patches are generated such that they appear achromatic when viewed in the D_{50} environment. Colorimetric measurements of those neutrals confirm that they have a chromaticity equal to that of D_{50}. The spectral characteristics that reproduced neutrals must have to achieve that result are discussed next.

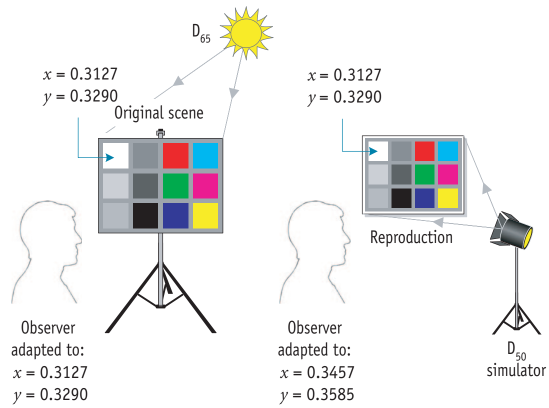


Figure 6.5 Under these viewing conditions, a colorimetric reproduction of a D_{65}-illuminated test-target neutral would not appear achromatic.

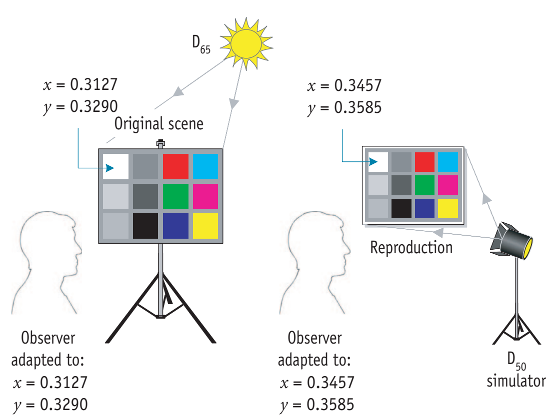


Figure 6.6 To appear achromatic under a D_{50} viewing illuminant, a reproduction of a test-target neutral must have a chromaticity equal to that of D_{50}.

Reflection neutrals: spectral characteristics

Neutrals, and other colors, are produced on three-dye reflection media from combinations of appropriate amounts of cyan, magenta, and yellow dyes. For any given set of CMY dyes, reflective support, and viewing illuminant, there is one and only one combination of dye amounts that will produce a visual neutral of a given luminance-factor value.

For example, Figure 6.7a shows the spectral characteristics for the dyes and reflective support of the photographic paper used in the experimental system. The dye characteristics are expressed in terms of spectral reflection density, which is the negative logarithm of spectral reflectance. Reflection density values, rather than reflectance values, are used here and throughout this discussion because changes in density values correlate more closely with changes in visual appearance. Figure 6.7b shows the relative amounts of those dyes required to form a neutral of visual reflection density 1.0 (a luminance-factor value $Y = 0.10$) when a D_{50} viewing illuminant is used. Also shown are the spectral reflection densities for that neutral and a spectrally nonselective neutral of 1.0 reflection density.

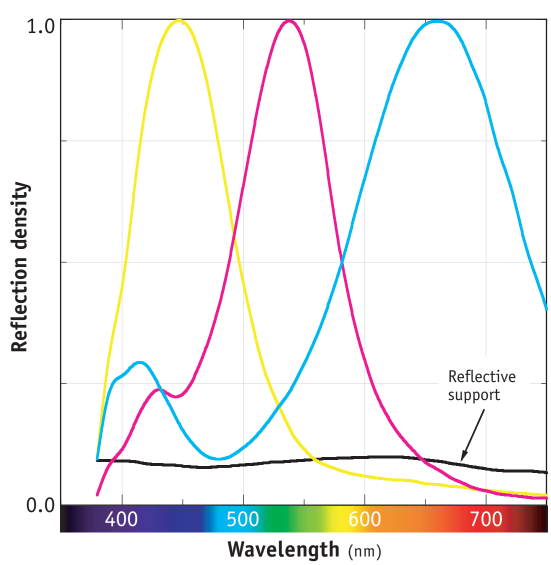


Figure 6.7a Spectral reflection density characteristics for a set of photographic cyan, magenta, and yellow dyes and a reflective support.

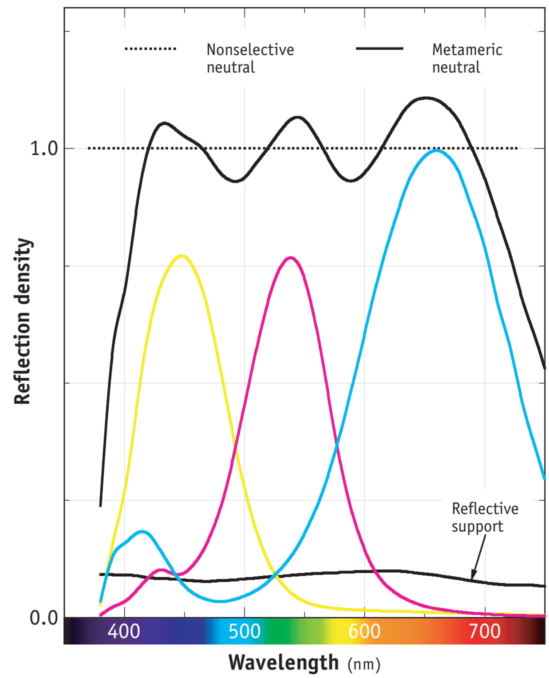


Figure 6.7b Spectral reflection densities for a nonselective neutral and a metamerically matching neutral formed using the cyan, magenta, and yellow dyes and reflective support of Figure 6.7a. Also shown are the relative amounts of the CMY dyes that, when added to the reflective support, form the metameric neutral.

Notice that the neutral formed by the CMY dyes is somewhat spectrally selective, i.e., it does not absorb or reflect light equally at each wavelength. The CMY neutral therefore is a *metameric* match to a spectrally nonselective neutral. As was discussed earlier, objects that are metameric matches under one illuminant may not match under another illuminant. Therefore, different spectral reflectances may be required in order to produce metameric neutral stimuli under different viewing illuminants. In a reflection color-imaging medium, spectral reflectances are changed by adjusting the amounts of the CMY dyes forming those reflectances.

Media in which the relative amounts of colorants must be changed significantly in order to form visual neutrals under different viewing illuminants are said to be viewing-illuminant sensitive. Reflection media can differ greatly in this characteristic, depending on the spectral absorption properties of their

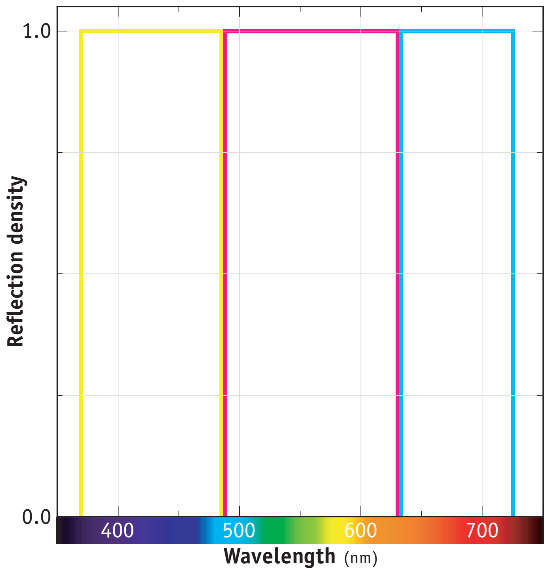


Figure 6.8a Spectral reflection density characteristics for a set of hypothetical cyan, magenta, and yellow block dyes.

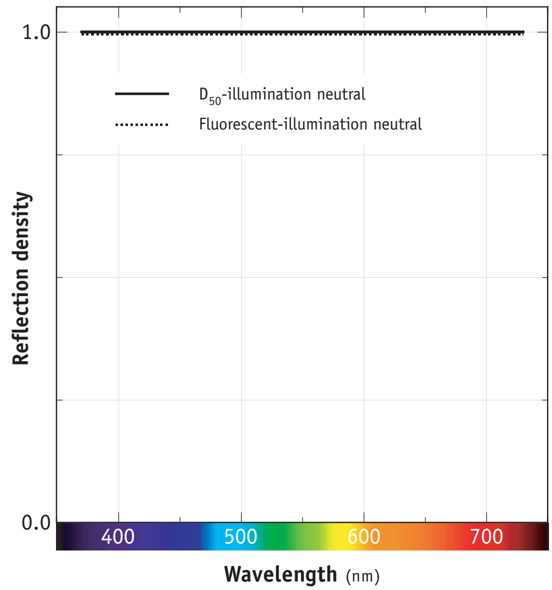


Figure 6.8b The neutrals produced by a reflection medium having block dyes are not viewing-illuminant sensitive. The spectral densities of the neutrals are identical for the two viewing conditions.

colorants. In the following figures, the degrees of viewing-illuminant sensitivity of the two reflection media of the experimental system and one hypothetical medium will be compared.

Figure 6.8a shows the spectral reflection densities for a set of CMY *block dyes* used in the hypothetical medium, and Figure 6.8b shows the spectral reflection densities for two neutrals formed by those dyes. The neutrals are those determined for two different viewing conditions. In one, the light source simulates CIE Standard Illuminant D_{50}; in the other, the light source is a particular type of fluorescent lamp. The spectral reflection densities of the two neutrals are *identical*, as would be expected. Since a neutral produced from these dyes absorbs (and reflects) light equally at each wavelength, it *always* will have the same chromaticity as the light source itself. If, as has been assumed, the observer chromatically adapts to the chromaticity of each light source, then an area that appears achromatic on this medium under one light source also will appear achromatic under any other light source. Neutrals formed on this medium therefore are not at all viewing-illuminant sensitive. That is a very desirable feature for a reflection medium to have, because a given reflection image is likely to be viewed under a variety of different illumination conditions.

The situation is somewhat different when the dyes of the photographic paper of the experimental system are used. The spectral characteristics of these dyes are shown in Figure 6.9a. Figure 6.9b shows that different spectral reflection densities are required in order to produce visual neutrals under each of the two previously described viewing conditions. However, although the spectral density characteristics of these neutrals are different, their basic similarities indicate that the photographic paper may not be highly viewing-illuminant sensitive. An area of this medium that appears achromatic when viewed under one light source likely will continue to appear achromatic, or at least nearly so, when viewed under most other light sources. In practice, that in fact is the case.

The spectral characteristics of the thermal transfer dyes are shown in Figure 6.10a. Figure 6.10b

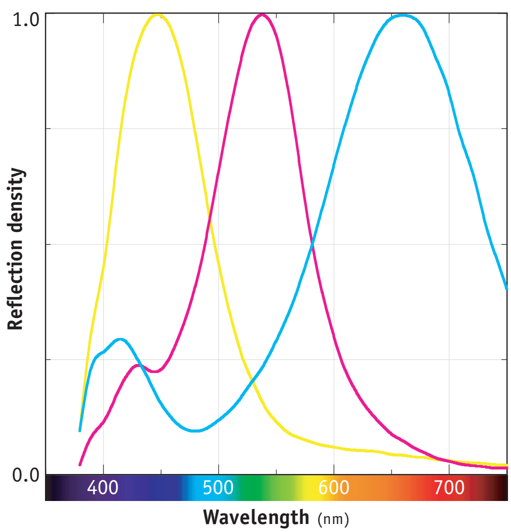


Figure 6.9a Spectral reflection density characteristics for a set of cyan, magenta, and yellow photographic paper dyes.

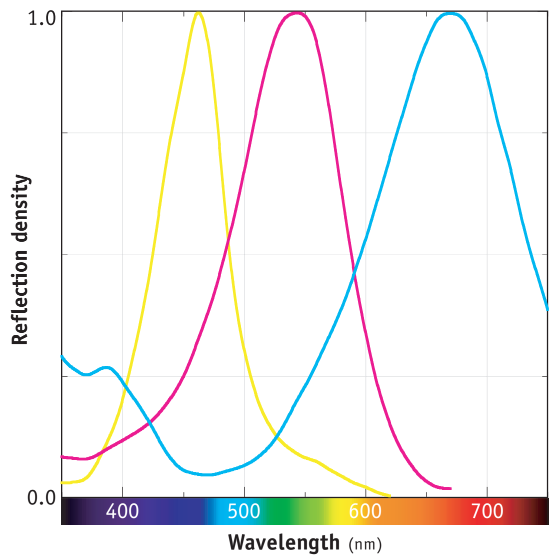


Figure 6.10a Spectral reflection density characteristics for a representative set of cyan, magenta, and yellow thermal transfer dyes.

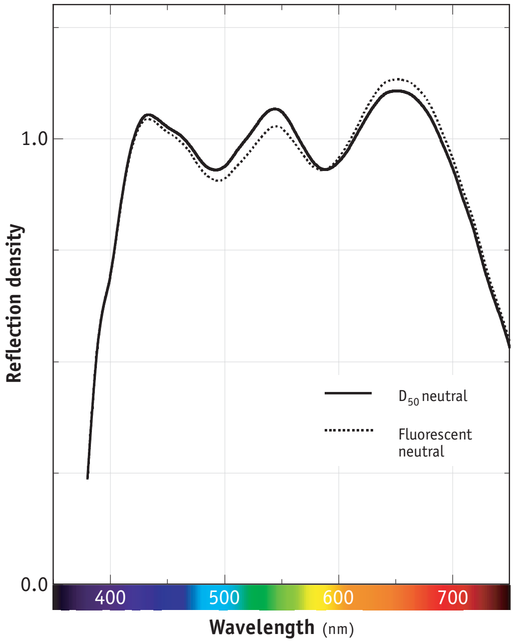


Figure 6.9b The neutrals produced by a reflection medium using the photographic paper dyes have relatively little viewing-illuminant sensitivity.

shows comparable spectral reflection densities for corresponding neutrals of the thermal dye-transfer medium used in the experimental system. Note that the neutrals of this medium are much more spectrally selective than those of the photographic paper. Moreover, note that quite different spectral reflection densities are required in order to produce neutrals under each of the two light sources. These are strong indications that the thermal medium is quite viewing-illuminant sensitive, which indeed it is.

The appearance of images produced on a viewing-illuminant-sensitive medium can be very different under different light sources. For example, if a typical thermal dye-transfer image is viewed under a variety of illumination conditions, such as by window light, fluorescent light, and tungsten light, the overall color balance of the image and the appearance of neutrals and other specific colors within the image will change significantly. For critical imaging applications, it may be necessary to produce custom images designed specifically for the particular illuminant that will be used for viewing.

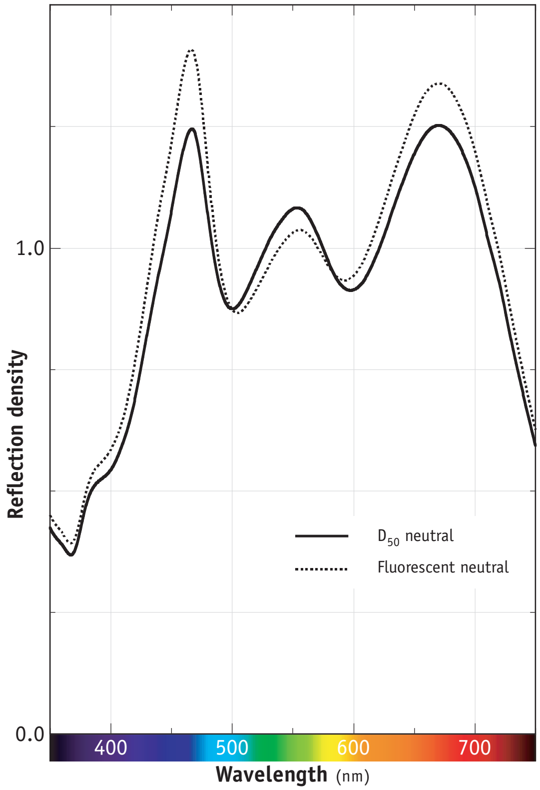


Figure 6.10b The neutrals produced by a reflection medium using the thermal transfer dyes of Figure 6.10a are viewing-illuminant sensitive.

Reflection neutrals: grayscale characteristics

The overall system grayscales—from the original scene to the reflection print—for each of the outputs of the experimental system can be characterized by photographing the test target and measuring the neutrals of the resulting prints. In this discussion, system grayscales will be specified in terms of reproduced visual densities (negative log luminance factor) as a function of original-scene log luminance factor. This is consistent with the grayscale analysis for the electronic imaging system of Chapter 5.

When the signal processing of the experimental system is adjusted such that high-quality reflection images of live scenes are obtained, the system grayscale characteristic (on either the photographic

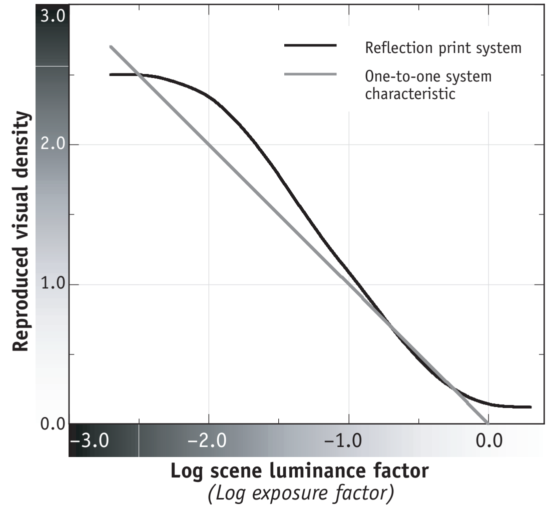


Figure 6.11 Grayscale characteristic for a high-quality reflection-print system, expressed in terms of reproduced visual density as a function of log scene luminance factor. A one-to-one system characteristic is shown for reference.

or the thermal medium) is substantially that shown in Figure 6.11. Similar results would be obtained from almost any high-quality reflection-print system used for reproducing live scenes, such as a system consisting of a color photographic negative film and color photographic paper.

The shape of the grayscale characteristic reveals that the system does not simply generate images that are one-to-one colorimetric reproductions of original scenes. As was the case for the electronic imaging system discussed previously, a one-to-one colorimetric reproduction of an original is appropriate only when the original *itself* is a reproduction, and only when the original and its reproduction are to be viewed under identical conditions.

Note that the grayscale characteristic for the reflection-print system is similar in many respects to that of an electronic system (Figure 4.22). That similarity is not unexpected, since many of the factors that required departures from one-to-one colorimetric reproduction for the electronic display are present in a displayed reflection-print display as well. One of those factors is viewing flare.

In reflection-image viewing, viewing flare occurs when incident light is reflected directly from the

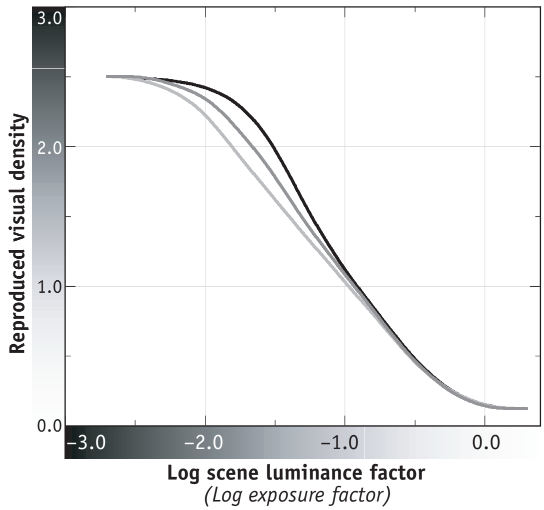


Figure 6.12 Reflection-print system grayscales incorporating compensation for three different levels of viewing flare. The black curve provides the most compensation of the three, the light-gray curve the least.

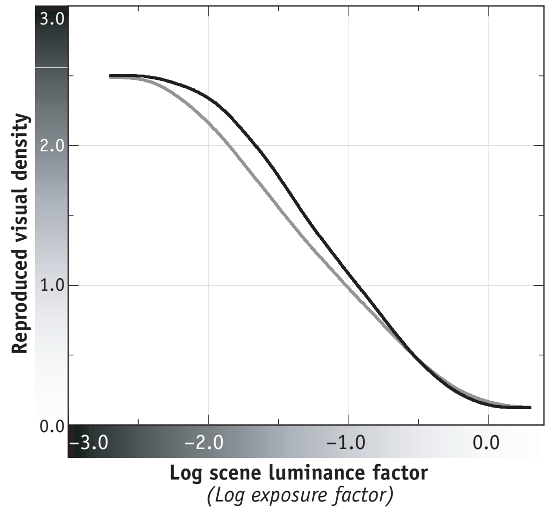


Figure 6.13 Increasing the photographic gamma of a system grayscale characteristic can provide compensation for a reduction in apparent luminance contrast. Both characteristics shown in the figure also include compensation for 0.75% viewing flare.

front surface of the medium. The amount of flare will vary, depending on the surface texture of the medium and other factors. As in the electronic system, the system grayscale must be designed such that it provides compensation for the anticipated amount of viewing flare.

Figure 6.12 shows three practical reflection-print system grayscales that incorporate compensation for different levels of viewing flare. Note that as more viewing-flare compensation is included, the higher-density regions of the grayscales are progressively increased in density. However, these densities cannot exceed the maximum density achievable by the medium. Therefore, the flare compensation becomes less successful as the required amount of that compensation is increased.

Another factor affecting the departure from a one-to-one grayscale reproduction is that reflection images generally are viewed indoors, where the level of illumination is much lower than is typical outdoors. The absolute luminances of images viewed indoors therefore will be relatively low. As discussed previously, images having lower absolute luminances will be perceived to have lower luminance contrast.

Compensation for this contrast effect can be achieved by increasing the overall *slope* of the grayscale characteristic (Figure 6.13). In photographic terms, that slope is referred to simply as gamma. However, to avoid confusion with the previously described gamma (the exponent γ used in CRT grayscale characteristic equations), the expression *photographic gamma* will be used from this point on when referring to measurements of hardcopy media.

Although the photographic gamma of any successful reflection-print system in which the originals are live scenes will be greater than unity, that of any given system will depend somewhat on its intended application. For example, reflection-print systems designed for portraiture generally are somewhat lower in photographic gamma, while those intended for amateur applications and some types of commercial advertising generally are higher. The grayscale characteristics of different reflection-print systems also may differ because of the particular capabilities and limitations of their reflection media. For example, each medium has a minimum density, which depends on the reflectance of its support material and the density of any overcoated layers.

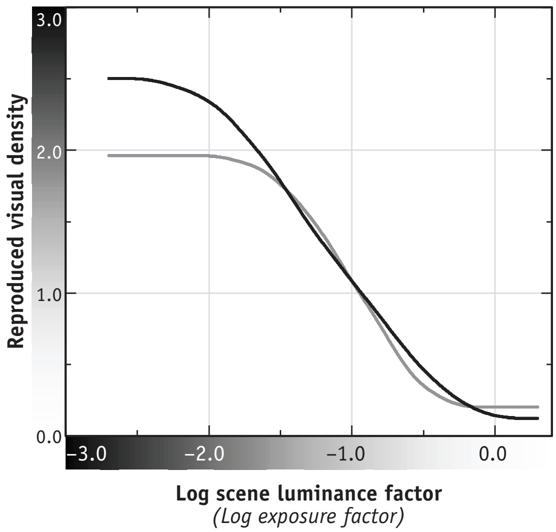


Figure 6.14 Grayscale characteristics for two reflection-print media having different minimum and maximum densities. The medium with the lower maximum density and higher minimum density (shown in gray) requires a higher photographic gamma in order to produce both whites and blacks at acceptable density levels.

Each medium also has a maximum density, which is limited by how much colorant can be formed or deposited in any given area and by certain other factors. The overall shape of the grayscale characteristic must be consistent with those density limits, as has been the case for all of the practical grayscale characteristics shown thus far.

In general, media having lower maximum densities and higher minimum densities must have somewhat higher mid-scale photographic gammas in order to produce whites and blacks at acceptable densities. This is illustrated in the grayscale characteristics of the two media shown in Figure 6.14.

Note that there is a significant amount of nonlinearity (beyond that required for flare compensation) in all of the reflection system grayscale characteristics that have been shown thus far. These nonlinearities are essential. They compress information from the highlight and shadow regions of the scene exposure range. That compression makes it possible for the reflection medium to display as much of the original exposure information as possible, given the minimum and maximum density limits of the medium itself.

Attempts sometimes are made to "correct" these grayscale nonlinearities by various means, particularly in digital imaging systems where grayscales can be linearized easily by the use of digital signal processing. However, whenever this is done, the resulting images inevitably are poor in quality because they fail to reproduce visually important details in highlight and shadow areas.

Reflection colors: colorimetric considerations

A meaningful evaluation of reproduced colorimetry from a color-imaging system must take into account effects due to the observer's state of chromatic adaptation. This is a particularly important consideration for reflection-print systems, because reflection images typically are viewed under conditions where there are numerous visual cues—such as white objects and specular reflections of light from the viewing illuminant—that cause the observer to chromatically adapt to the conditions of the viewing environment rather than to the color-reproduction properties of the images themselves.

When the observer's state of chromatic adaptation differs between the original-scene and reproduced-image viewing environments, reproduced colorimetric values should be compared to original colorimetric values that have first been *transformed* appropriately to visually equivalent colorimetric values. The transformed values describe a color stimulus that would produce, for an observer chromatically adapted to the reproduced-image viewing illuminant, a visual match to the original color stimulus viewed by an observer who is chromatically adapted to the original-scene illuminant.

That last sentence may warrant rereading; the idea is not easy to grasp. But it is critical that the concept of chromatic adaptation be understood before proceeding. The appropriate use of chromatic adaptation transforms is essential for evaluating color-imaging systems, as is being done here. In addition, such transforms play a major role in the appearance-based color-encoding methods that will be discussed later.

Table 6.1 Comparison of the original test target XYZ tristimulus values for CIE D65 illumination and visually equivalent XYZ tristimulus values for an observer chromatically adapted to the chromaticity of CIE D_{50}. Equivalent tristimulus values were computed using a von Kries chromatic adaptation transform.

	D_6-Illuminated Testy-Target Values			Visually Equivalent *XYZ* Values for a D_{50}-Adapted Observer		
Test-Target Color	X_{D65}	Y_{D65}	Z_{D65}	D_{D50}	Y_{D50}	Z_{D50}
Neutral 1	0.9504	1.0000	1.0889	0.9642	1.0000	0.8249
Neutral 2	0.5598	0.5890	0.6414	0.5679	0.5890	0.4859
Neutral 3	0.3450	0.3630	0.3953	0.3500	0.3630	0.2995
Neutral 4	0.1901	0.2000	0.2178	0.1929	0.2000	0.1650
Neutral 5	0.0846	0.0890	0.0969	0.0858	0.0890	0.0734
Neutral 6	0.0304	0.0320	0.0348	0.0308	0.0320	0.0264
Red	0.2000	0.1180	0.0560	0.2068	0.1186	0.0424
Green	0.1480	0.2380	0.0990	0.1584	0.2377	0.0750
Blue	0.0850	0.0620	0.3070	0.0738	0.0619	0.2326
Cyan	0.1490	0.2050	0.4040	0.1416	0.2045	0.3061
Magenta	0.2980	0.1970	0.3240	0.2968	0.1975	0.2455
Yellow	0.6270	0.6820	0.0980	0.6697	0.6826	0.0742

A number of different transformation techniques can be used to determine visually equivalent colorimetric values. At present, the CIE is seeking an agreed-upon model of color vision that will include a process for performing highly accurate chromatic adaptation transformations. Of the existing techniques, a von Kries matrix transformation, which is described in Appendix D, is one of the simplest. The method is used widely, and it generally works quite well, especially in situations where the differences in adaptation chromaticities are relatively small.

The matrix equation below is an example of a von Kries transformation of CIE *XYZ* values for D_{65} chromatic adaptation to visually equivalent CIE *XYZ* values for D_{50} chromatic adaptation:

$$\begin{bmatrix} X_{D_{50}} \\ Y_{D_{50}} \\ Z_{D_{50}} \end{bmatrix} = \begin{bmatrix} 1.0161 & 0.0553 & -0.0522 \\ 0.0060 & 0.9956 & -0.0012 \\ 0.0000 & 0.0000 & 0.7576 \end{bmatrix} \begin{bmatrix} X_{D_{65}} \\ Y_{D_{65}} \\ Z_{D_{65}} \end{bmatrix} \tag{6.1}$$

The $XYZ_{D_{50}}$ values describe a color stimulus that would produce, for an observer chromatically adapted to the chromaticity of D_{50}, a visual match to the original $XYZ_{D_{65}}$ stimulus viewed by an observer who is chromatically adapted to the chromaticity of D_{65}. Note that if the *XYZ* tristimulus values for CIE Standard Illuminant D_{65} ($X = 95.04$, $Y = 100.00$, $Z = 108.89$) are transformed using this matrix, the resulting *XYZ* tristimulus values ($X = 96.42$, $Y = 100.00$, $Z = 82.49$) will be those for CIE Standard Illuminant D_{50}. This is consistent with the assumption that the observer is chromatically adapted to the chromaticity of the viewing illuminant.

Table 6.1 lists the original and transformed colorimetric values for each of the patches of the test target. The chromaticity coordinates for the two sets of tristimulus values are compared graphically in Figure 6.15. The chromaticities are plotted in terms of CIE u', v' coordinates, rather than CIE x, y coordinates, to better illustrate the magnitudes of the colorimetric transformations.

Reflection colors: other characteristics

In the experimental system, grayscale modifications required to compensate for viewing flare and low

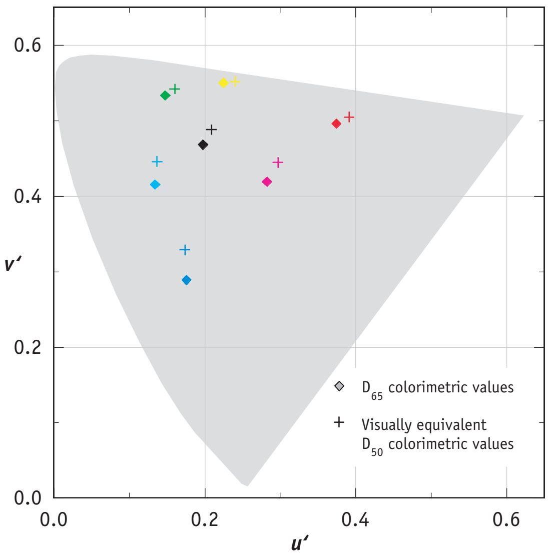


Figure 6.15 CIE u', v' diagram showing a comparison of test-target chromaticities for D_{65} illumination (◇) and visually equivalent chromaticities based on D_{50} chromatic adaptation (+). The equivalent values were computed using a von Kries transformation matrix.

absolute luminances are accomplished by adjusting *RGB*, rather than luminance, signal processing. Doing so has the effect of altering the characteristics of the cyan, magenta, and yellow scales that make up the grayscale (Figure 6.16). This results in higher overall reproduced color saturation, which is desirable because it helps compensate for color-saturation decreases associated with viewing flare and with the relatively low absolute luminance levels of the reflection images (Figure 6.17).

In most reflection-print systems, additional signal processing is used to further modify color reproduction. Color reproduction may be altered to compensate for color-memory effects, to account for various color preferences, or to adjust the overall color saturation to best meet the goals of the intended application. The exact methods used by manufacturers to adjust and optimize the color reproduction of their media and systems vary, and such methods usually are proprietary. However, the general ideas behind these methods, and their relationship to the concepts that have been discussed, can be illustrated for the experimental system.

Figure 6.18 shows that in the experimental system, the electronic camera has spectral responsivities equivalent to a set of all-positive color-matching functions. A matrix is used to transform the linear *RGB* color-signal values produced by the camera to CIE $XYZ_{D_{65}}$ values. This is followed by a chromatic adaptation matrix to transform the D_{65} colorimetric values to visually equivalent D_{50} colorimetric values. A third matrix, which is a function of the particular cyan, magenta, and yellow dye set used in the output medium, then is used to transform from $XYZ_{D_{50}}$ values to RGB_m values.

Appropriate nonlinear processing then is used such that the overall system grayscale characteristic corresponds to that shown in Figure 6.11. Finally, a fourth matrix, or some other form of three-dimensional signal processing, is used to create desirable interactions among the nonlinear color signals. One factor that contributes to the need for these interactions is discussed below and illustrated in Figures 6.19a, 6.19b, and 6.19c.

Figure 6.19a shows spectral density curves of a representative set of cyan-, magenta-, and yellow-image-forming dyes. These spectra were used to derive an example series of corresponding red, green, and blue primary spectral transmission curves by subtracting sequentially increasing amounts of each of the dye curves from a perfect spectral black (Figure 6.19b). The red spectral transmission series in the figure, for instance, was formed by subtracting various amounts of the cyan dye from black. The resulting red, green, and blue spectral transmission series can be compared to additive-primary red, green, and blue intensity series, such as those shown previously in Figure 4.6a. While the wavelength-to-wavelength ratios of each additive-primary spectrum are constant throughout an intensity series, it is apparent from Figure 6.19b that the wavelength-to-wavelength ratios of each subtractive-primary spectrum vary throughout a comparable transmission series. This would be expected. For a dye (on a clear support), the wavelength-to-wavelength ratios are constant in *density* (logarithmic) space and not in linear (transmission) space. Because the linear ratios are changing, the chromaticity values of each subtractive primary are not constant throughout the

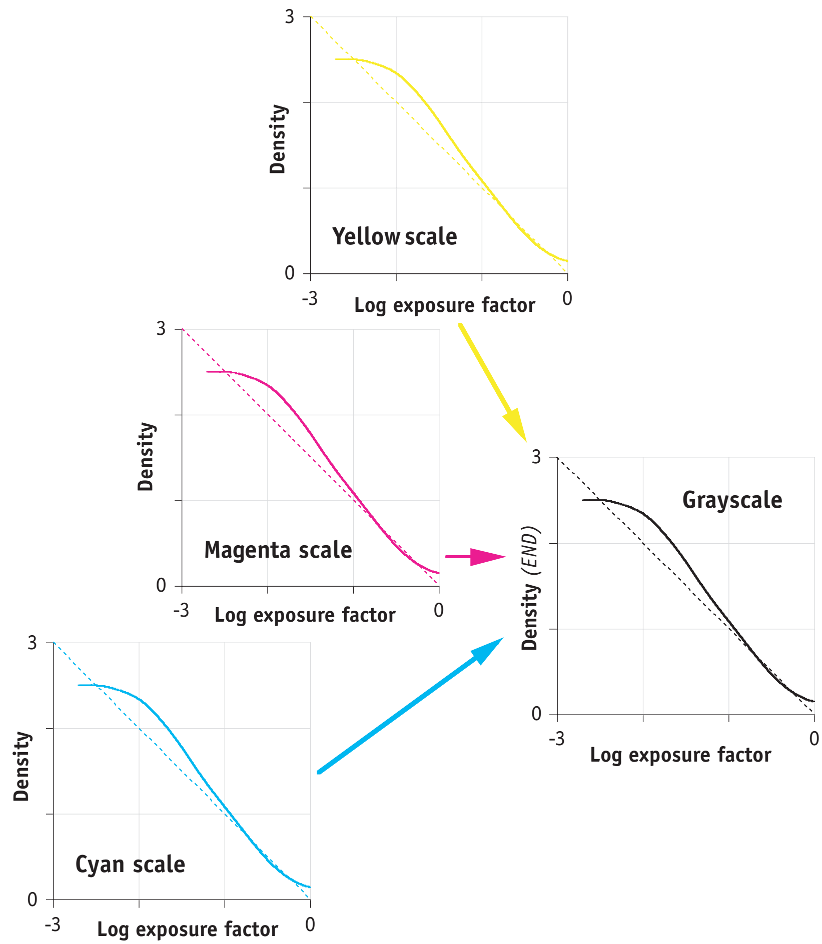


Figure 6.16 A subtractive grayscale and its component cyan, magenta, and yellow dye scales.

series, as illustrated in Figure 6.19c. Similar chromaticity shifts also occur when subtractive primaries are used on reflective supports.

This lack of "chromatic stability," inherent in subtractive primaries, requires appropriate compensation in the form of nonlinear chemical and/or electronic three-dimensional interactions. In addition, nonlinear signal processing interactions also are required in subtractive systems in order to properly adjust overall color saturation and to produce other desirable colorimetric alterations. A discussion of how these interactions are optimized is well beyond the scope of this book. For our purposes, it is most important to recognize that such interactions contribute to the significant color-reproduction differences that exist among various reflection-print media and systems. Those differences will be addressed later in our examination of color-management systems.

Reflection system colorimetry

When appropriate signal processing is used, a reflection-print system based on the experimental system will produce excellent-quality images. The colorimetric results of the overall system are shown in Figure 6.20. These results show that the reproduced colorimetry matches the modified

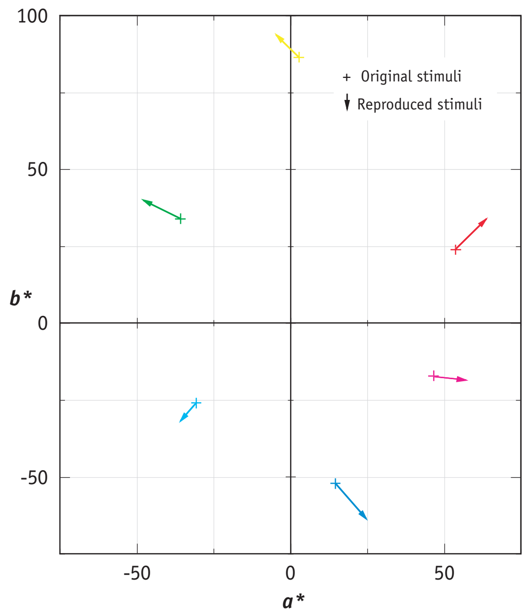


Figure 6.17 The vector arrows indicate the chroma increases associated with the photographic gamma increase and viewing-flare compensation described in the text. The a^*b^* values are based on measurements made in the absence of flare light.

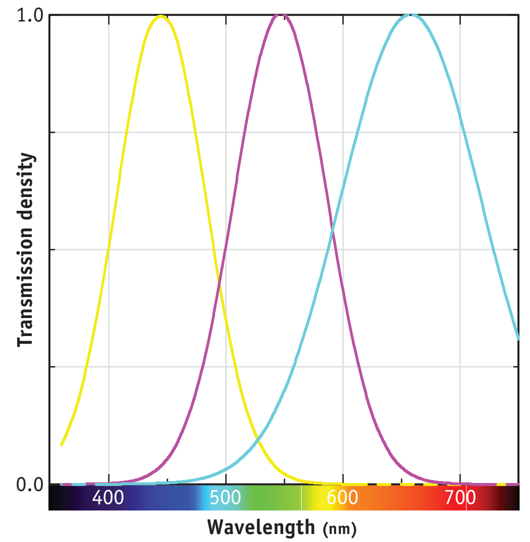


Figure 6.19a Spectral density spectra for a set of cyan-, magenta-, and yellow-image-forming dyes.

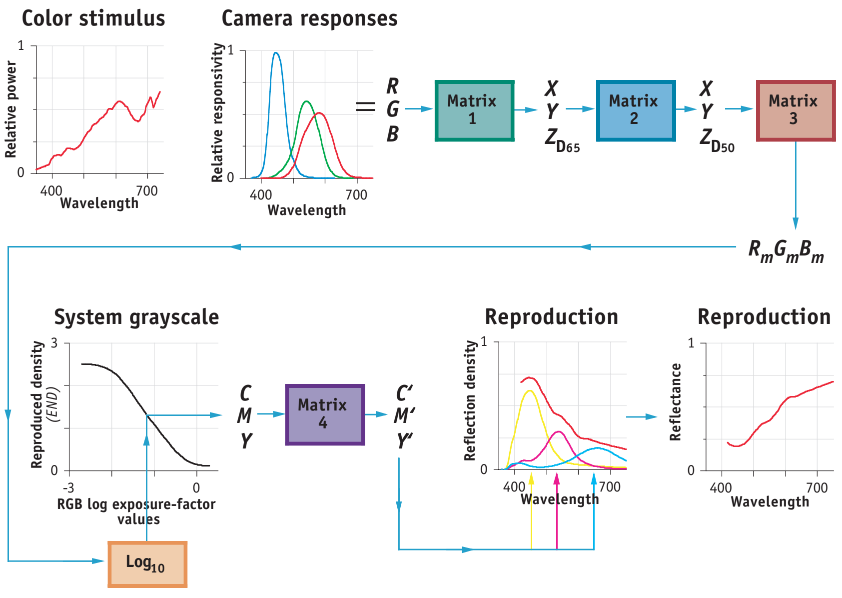


Figure 6.18 Signal processing in the experimental reflection-print color-imaging system. The output is shown in terms of both reflection density and linear reflectance.

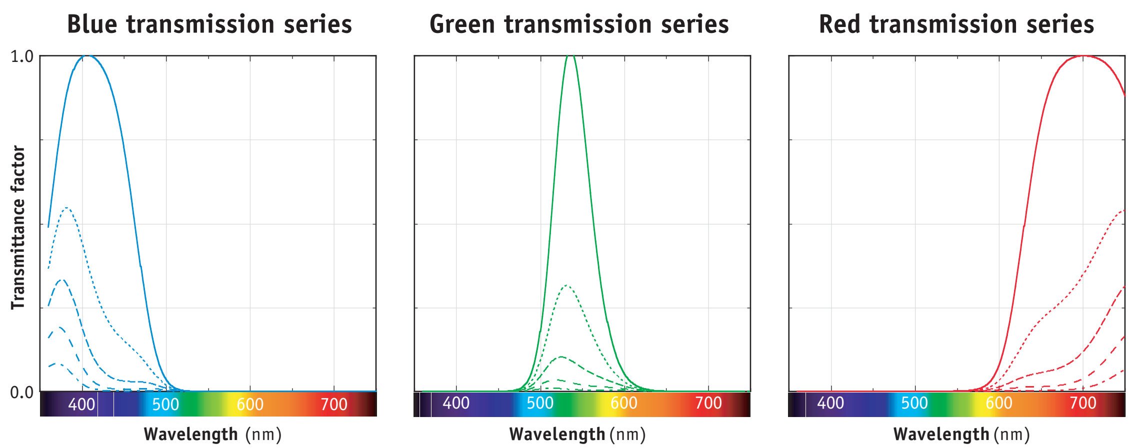


Figure 6.19b A set of red, green, and blue spectral transmission series derived using the dyes of Figure 6.19a. Note that the wavelength-to-wavelength transmission ratios of each primary differ at each transmission level.

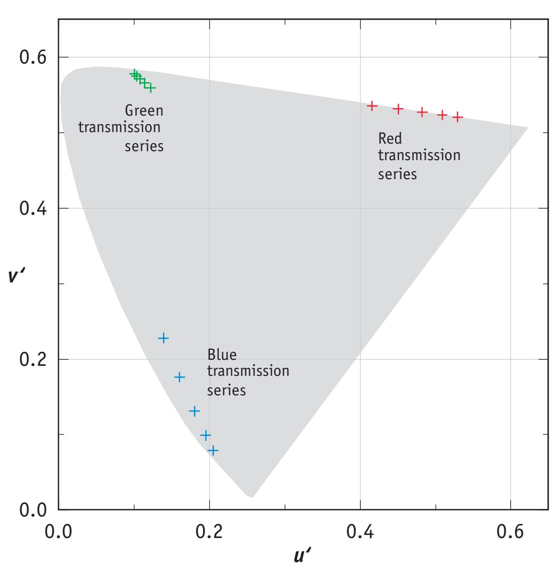


Figure 6.19c Chromaticities for the spectral transmission series of Figure 6.19b. The chromaticity of each primary does not remain constant throughout a transmission series.

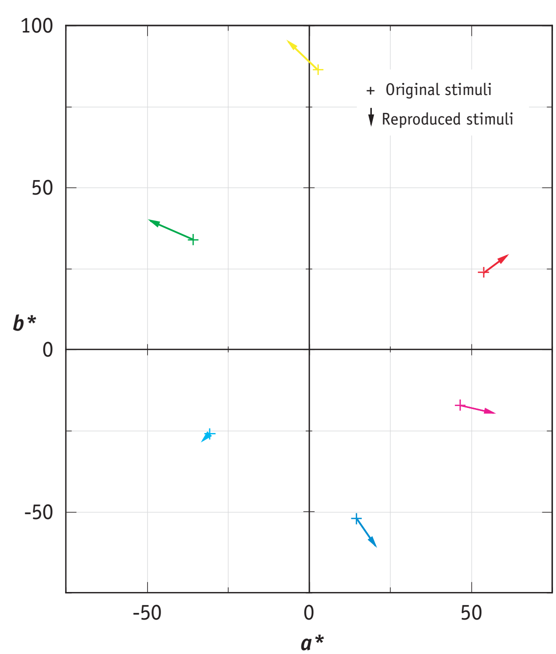


Figure 6.20 Colorimetric results of an actual reflection-print color-imaging system based on the experimental system. The + marks indicate equivalent (chromatically adapted) D_{50} values for the D_{65}-illuminated test target. The vector arrowheads indicate the overall system reproductions.

colorimetry of Figure 6.17 for all colors within the gamut of the output medium. These results also show once again that high-quality reproductions of original scenes are not simply one-to-one reproductions of the original colorimetry.

The colorimetric behavior of other actual reflection-print systems corresponds to that of this experimental system to various degrees. As mentioned earlier, some differences are deliberate and are based on the specific application for which the system is designed. Other differences are unavoidable, due to limitations of the particular devices and media that are involved. For example, implementing all of the color signal processing that has been described for the experimental system can be quite difficult in a conventional photographic system where that processing must be accomplished using chemistry alone.

Color-encoding considerations

This analysis has raised several issues that have consequences related to the scanning and color encoding of reflection images. The fact that reflection images are objects that produce stimuli only when illuminated means that the colorimetry they ultimately will produce cannot be determined and numerically represented unless the spectral power distribution of the intended viewing illuminant is specified. Meaningful interpretation of the resulting colorimetric values requires that the amount of flare in the viewing environment and the observer's state of chromatic adaptation also be known.

Viewing-illuminant sensitivity was discussed in terms of changes in image appearance with changes in viewing illuminant, but such sensitivity also is an important consideration for input scanning. Viewing-illuminant-sensitive media are problematic input-image sources. The color values that are measured from such media are strongly affected by the particular spectral properties of the scanner, such as the spectral power distribution of its illuminant and the spectral sensitivities of its sensor.

This analysis also has described some of the reasons why different reflection-print systems and media have unique and distinctive appearances. Whether such distinctions should be retained fully, modified, or removed entirely during the color-encoding process is an important issue that will need to be revisited in later discussions.

Summary of key issues

- Reflection images are objects that produce stimuli only when illuminated. Their colorimetry is directly affected by the spectral characteristics of the illuminant.
- Reflection media differ greatly in their sensitivity to changes in viewing illuminant spectral power distribution.
- Reflection images are perceived as objects within a viewing environment. The observer's state of adaptation is determined by that environment. There is little or no adaptation to the image itself.
- Differences in the observer's adaptive state between the original and print viewing environments must be accounted for in the design of a reflection-print system.
- A chromatic adaptation transformation can be used to determine visually equivalent tristimulus values for two states of observer chromatic adaptation.
- An optimum reflection-print system must depart from a one-to-one colorimetric relationship with the original scene in order to compensate appropriately for viewing flare and to provide desirable colorimetric alterations.
- Different reflection-print systems and media have distinctive appearances that result from their particular capabilities and limitations and from differences in the applications for which they were designed.

7

Projected Images

Projected images can be created using any of a variety of image display technologies, devices, and media. Examples include optical projection of photographic slide films, optical projection of motion picture films, and electronic projection using laser, CCD, or DLP technologies. Yet despite differences in the underlying technologies that might be used to generate them, virtually all forms of projected images share certain fundamental colorimetric characteristics and relationships to original scenes. Based on the discussions of the preceding chapters, it is not unexpected that the colorimetry of projected images is very different from that of original scenes. What might not be expected is that the colorimetry of projected images also differs significantly from that of the electronic display and reflection-print systems described so far.

In this chapter, the colorimetric characteristics particular to projected images will be analyzed, and the reasons why those characteristics must differ from those of other forms of image display will be explained. The discussion will begin with a study of the colorimetric properties of photographic slide films—in part because of their historical importance, but more importantly because their fundamental colorimetric properties are representative of those of all projected images. The colorimetric properties of other projection media and devices and other related modes of display also will be shown. Finally, several important digital color-encoding considerations related to the colorimetric properties of these forms of image display will be discussed.

Photographic slide films

Until quite recently, photographic transparency films such as 35 mm slide films and large-format sheet films were the most commonly used form of input for high-end color-imaging systems. There were good reasons for this: photographic transparency films are capable of extraordinary image quality—high sharpness, low noise—and they can record and reproduce a wide range of colors. In addition, they can be viewed directly for evaluation and selection, which was an important consideration to editors of catalogs, magazines, and other publications prior to the widespread use of high-speed softcopy previewing and proofing.

Photographic slide films: general characteristics

Figure 7.1 shows a simplified cross-section of a photographic slide film. The film has a blue-light-sensitive layer, a green-light-sensitive layer, and a red-light-sensitive layer, which are coated on a transparent support. Because both the green-light-sensitive and red-light-sensitive layers also are inherently sensitive to blue light, a yellow filter layer is coated above

Digital Color Management: Encoding Solutions, Second Edition Edward J. Giorgianni and Thomas E. Madden
 ISBN: 978-0-470-51244-9

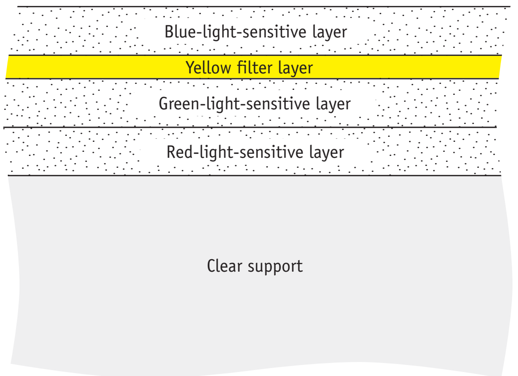


Figure 7.1 Simplified cross-section of a photographic transparency film (prior to chemical processing).

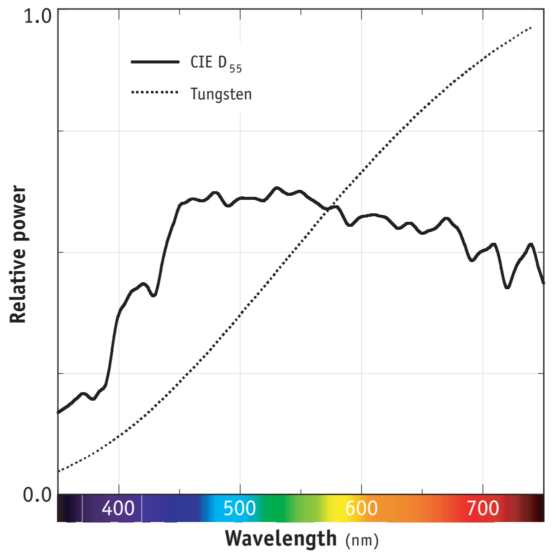


Figure 7.2 Spectral power distributions of CIE Standard Illuminant D_{55} (solid line) and a tungsten light source (dotted line).

these layers to prevent any blue light from reaching them.

Photographic slide films are designed to be used with specific image-capture illuminants. For example, some slide films are designed for daylight-illumination photography, while others are designed for tungsten-illumination photography. In the manufacturing of the film, the relative sensitivities (*speeds*) of the red-light-sensitive, green-light-sensitive, and blue-light-sensitive layers are balanced such that properly color-balanced images result when scenes illuminated by the reference scene illuminant are photographed. Use of any other type of scene illuminant most likely will result in images having an overall color-balance shift.

For example, Figure 7.2 shows the relative spectral power distributions for CIE Standard Illuminant D_{55} and a tungsten light source. Figure 7.3 shows the red, green, and blue spectral sensitivities of a representative *daylight* transparency film, which is balanced for D_{55} scene illumination. If this film were used to photograph a scene illuminated by a *tungsten* light source, the resulting slides would have an overall orange (red-yellow) color balance due to the relatively higher red power and lower blue power of the tungsten source.

Figure 7.3 also shows the red, green, and blue spectral sensitivities for a representative photo-

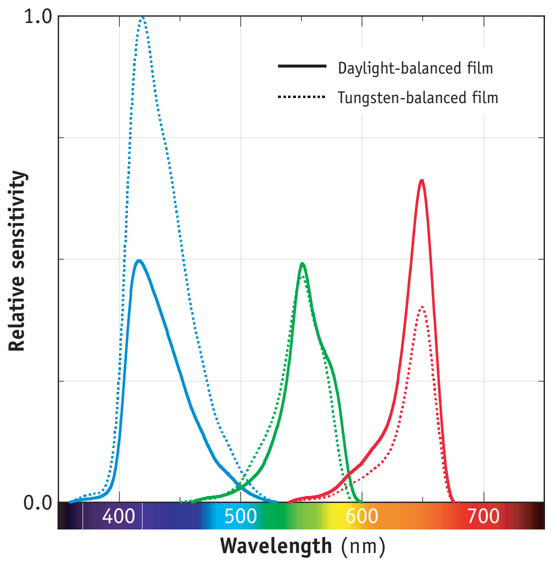


Figure 7.3 Red, green, and blue spectral sensitivities of representative daylight-balanced and tungsten-balanced photographic transparency films.

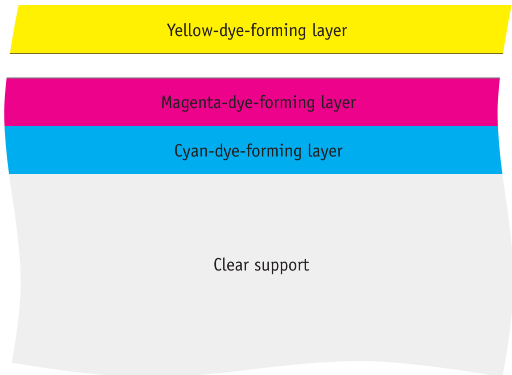


Figure 7.4 Simplified cross-section of a photographic transparency film (after chemical processing). Yellow, magenta, and cyan image dyes are formed in the blue-light-sensitive, green-light-sensitive, and red-light-sensitive layers, respectively.

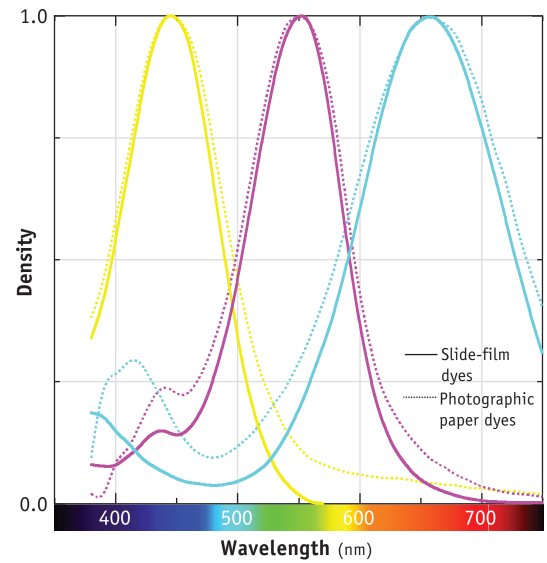


Figure 7.5 Spectral densities of the cyan, magenta, and yellow image-forming dyes of a representative photographic slide film (solid lines) compared to those of a representative photographic paper (dotted lines).

graphic transparency film that is balanced specifically for tungsten-illumination photography. Note that its red sensitivity is relatively low, and its blue sensitivity is relatively high, compared to the respective sensitivities of the daylight film.

When an exposed area of a photographic transparency film is chemically processed, yellow, magenta, and cyan image dyes are formed in the blue-light, green-light, and red-light-sensitive layers, respectively (Figure 7.4). A *positive* image results, i.e., the *maximum* amount of dye forms at the *minimum* exposure; the *minimum* amount of dye forms at the *maximum* exposure. Also during chemical processing, the yellow filter layer is made colorless. (Refer to Appendix C for more details on photographic media.)

Figure 7.5 shows the spectral transmission densities of the cyan, magenta, and yellow image-forming dyes of a representative photographic transparency film. Note that these dyes seem to be purer than those of the photographic paper, which are also shown in Figure 7.5. *Purer* means that each dye more selectively absorbs light of just one primary color, and each has less unwanted absorption of light of the other two primary colors. For example, the cyan dye of the photographic transparency film absorbs mostly red light and relatively little green or blue light. In addition to absorbing red light, however, the cyan dye of the photographic paper absorbs significantly more green and blue light.

This apparent difference in purity is not due to differences in the inherent qualities of the dyes that are used. It is due to the fundamentally different optical characteristics of transmission and reflection media. The internal reflections of light within a reflective-support medium, previously shown (Figure 6.3), effectively increase the unwanted spectral absorptions of its image-forming dyes.

For example, a dye that appears to be a pure yellow on a transmissive support may appear quite orange on a reflective support because the reflection optics magnify any unwanted green-light absorption the dye might have. This is shown in Figures 7.6a and 7.6b. In Figure 7.6a, the spectral transmission–density characteristics and the spectral reflection–density characteristics for the same yellow dye are compared. Note that the dye exhibits

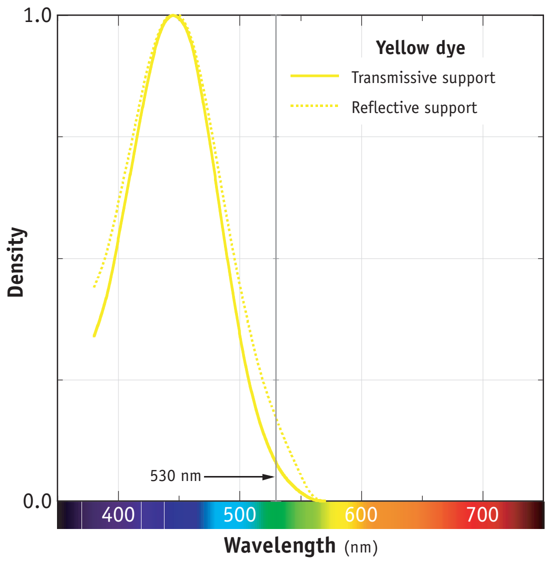


Figure 7.6a Spectral transmission density and spectral reflection density for the same yellow dye. The dye amounts used are adjusted such that the peak reflection densities and peak transmission densities are equal.

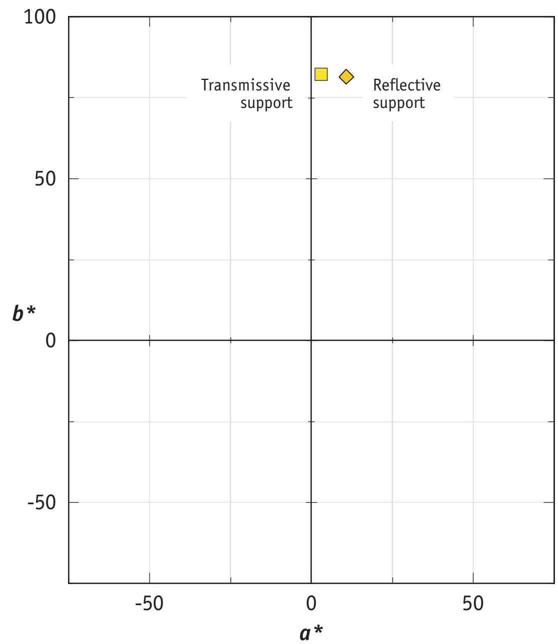


Figure 7.6b Hue change, from yellow toward orange, which results when the same yellow dye is used on a transmissive support and on a reflective support.

greater unwanted absorption of green light (at 530 nm, for example) when used on a reflective support. Figure 7.6b shows the effective hue change of the yellow dye. The vector in the figure exhibits a clockwise rotation, which corresponds to a move from a purer yellow to a redder (more orange) yellow.

As a result of this effect, when all other factors are equal, a transmission medium will have a larger color gamut than that of a reflection medium having the same CMY image-forming dyes (Figure 7.7). Also as a result of this effect, again when all other factors are equal, a color-imaging system using a transmission medium for output will require less signal processing (color correction) than an otherwise comparable system using a reflection output medium.

Photographic slide films: neutral characteristics

Like reflection images, images on photographic transparency media are objects that produce color stimuli only when illuminated. The colorimetry of a transparency film neutral, therefore, will be affected directly by the spectral power characteristics of the viewing illuminant.

Figure 7.8 shows the spectral characteristics for a colorimetric neutral produced by the CMY image-forming dyes of a representative photographic transparency film. The neutral, which has a visual density of 1.00, was determined for a particular viewing illuminant. Note that the neutral is quite spectrally selective. This suggests that it will be viewing-illuminant sensitive, which indeed is the case.

A comparable degree of viewing-illuminant sensitivity would be a very serious problem for a reflection medium, because reflection images are likely to be viewed under a wide variety of conditions. However, a given transparency film, in addition to being designed for a specific reference *scene* (image-capture) illuminant, also is designed for a specific reference *viewing* illuminant.

For example, 35 mm slide films are designed to be projected using a particular type of

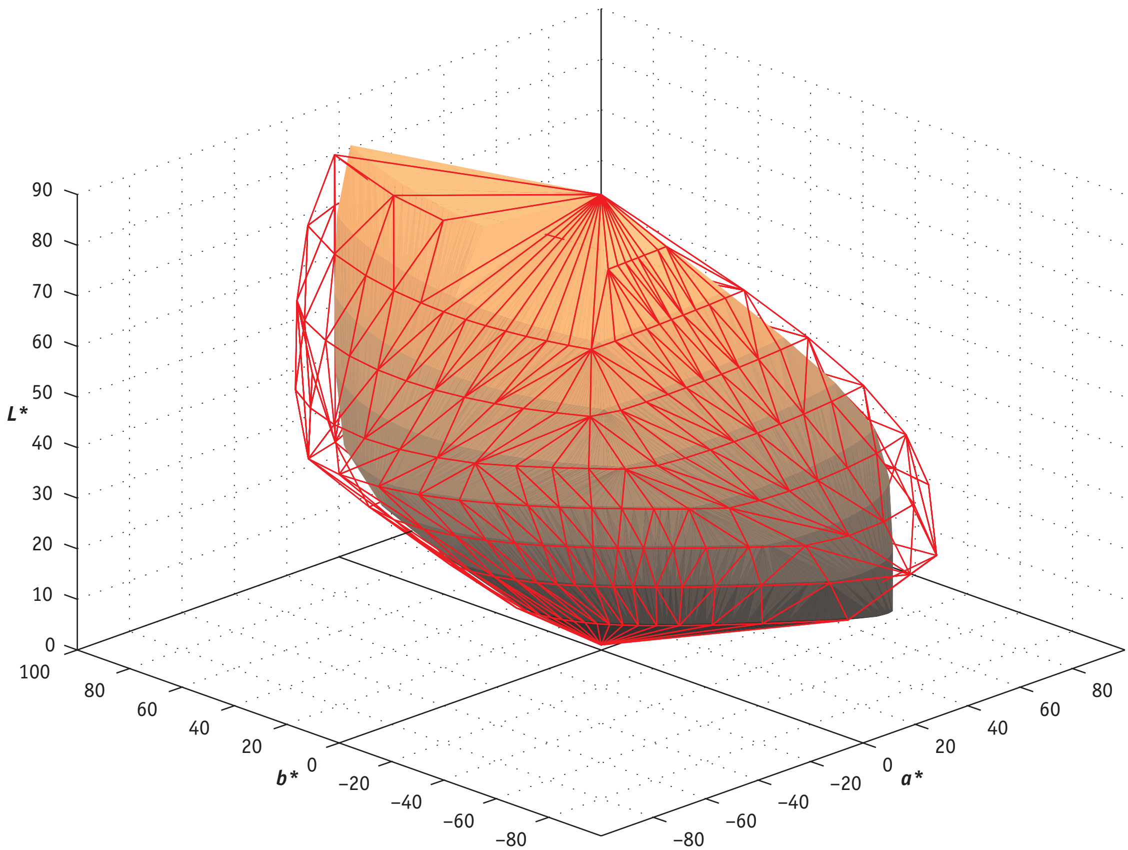


Figure 7.7 Comparison of the resulting color gamut when the same set of CMY image-forming dyes is used on a transmissive support (the gamut boundary shown by the wire frame) and a reflective support (gamut boundary shown by the solid).

tungsten–halogen projection light source. Larger sheet films are designed to be viewed on a backlight box having an artificial D_{50} light source. In the engineering of most transparency films, therefore, viewing-illuminant sensitivity is of lesser concern, and it generally is traded off for other features. In particular, CMY image-forming dyes having a larger color gamut can be used when viewing-illuminant sensitivity is not a major concern.

This type of trade-off is illustrated in Figures 7.9a, 7.9b, and 7.9c. These figures show that although the viewing-illuminant sensitivity of dye Set *A* is greater than that of Set *B*, Set *A* also has lesser amounts of unwanted absorptions. Thus Set *A* has a somewhat larger color gamut, and it requires less color correction to produce equivalent colors.

Photographic slide films: color balance

The neutral previously shown in Figure 7.8 is a colorimetric neutral, i.e., the CIE x, y chromaticity coordinates of that neutral, when projected, are identical to those of the projection light source itself. In Figure 7.10, the colorimetric neutral is compared to an *actual* neutral produced by a representative commercially successful photographic slide film having the same set of CMY dyes. The actual neutral is a reproduction of a spectrally nonselective neutral patch of the test target, which was photographed using the reference scene illuminant.

The figure would seem to suggest that the actual neutral produced by the slide film is cyan-blue,

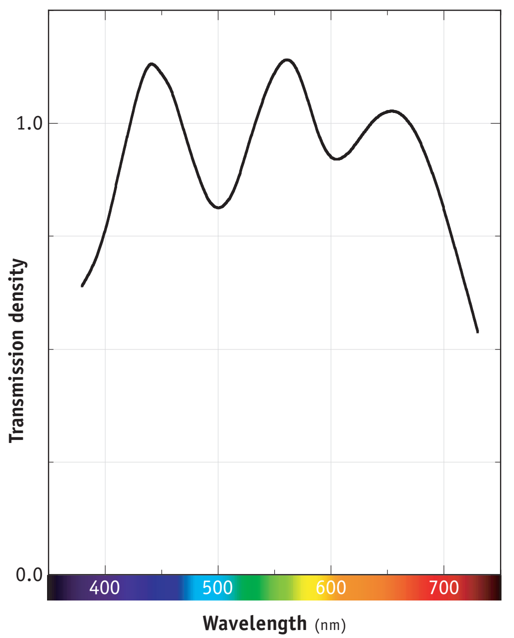


Figure 7.8 Colorimetric neutral produced by the CMY image-forming dyes of a representative photographic transparency film. The viewing illuminant is a tungsten–halogen projection lamp.

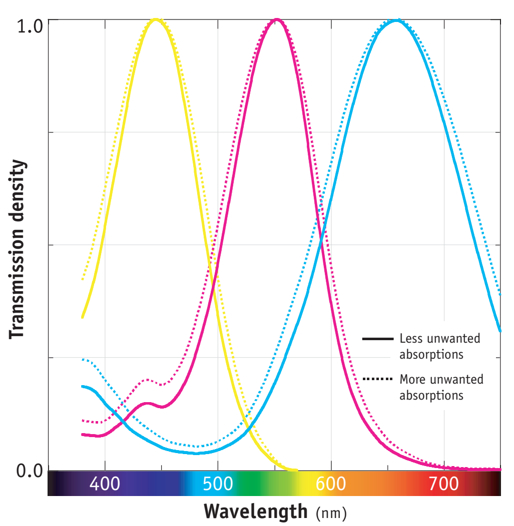


Figure 7.9a Spectral transmission density characteristics of two sets of cyan, magenta, and yellow image-forming dyes. The dyes of Set *A* (solid lines) have much lesser amounts of unwanted absorptions than do the dyes of Set *B* (dotted lines).

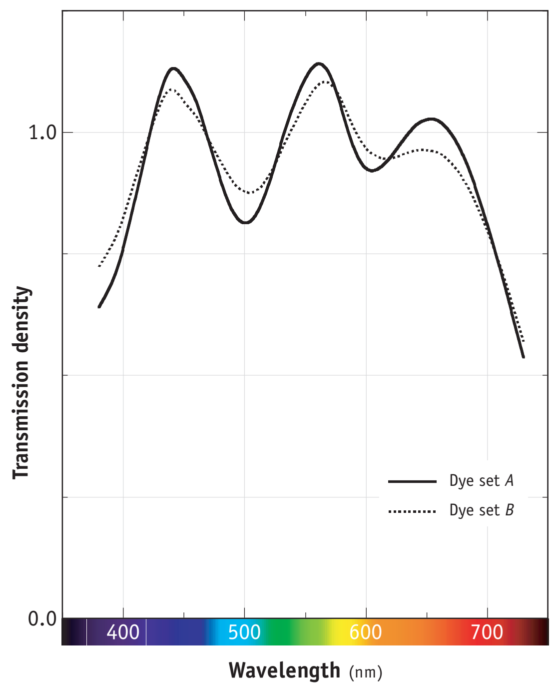


Figure 7.9b Neutral formed by the dyes of Set *A* are more spectrally selective—and therefore more viewing-illuminant sensitive—than those formed by the dyes of Set *B*.

since it has relatively more density to red light (more cyan dye) and relatively less density to blue light (less yellow dye). In fact, density measurements, colorimetric measurements, dye-amount measurements, or any other physical measurements that could be made, all would indicate that the actual film neutral is indeed cyan-blue. The only disagreement on this point would come from one source: human observers.

When observers view the projected "cyan-blue" actual film neutral, they judge it to be neutral (achromatic). And when observers view the projected *colorimetric* neutral, they judge it to be somewhat reddish yellow. Why? Since other visual cues are absent in a darkened room, an observer will tend to adapt to a projected image itself. If that adaptation were *complete*, then *either* projected neutral, viewed individually, would appear achromatic.

However, the light emitted by a tungsten–halogen projection lamp is quite reddish yellow, the absolute

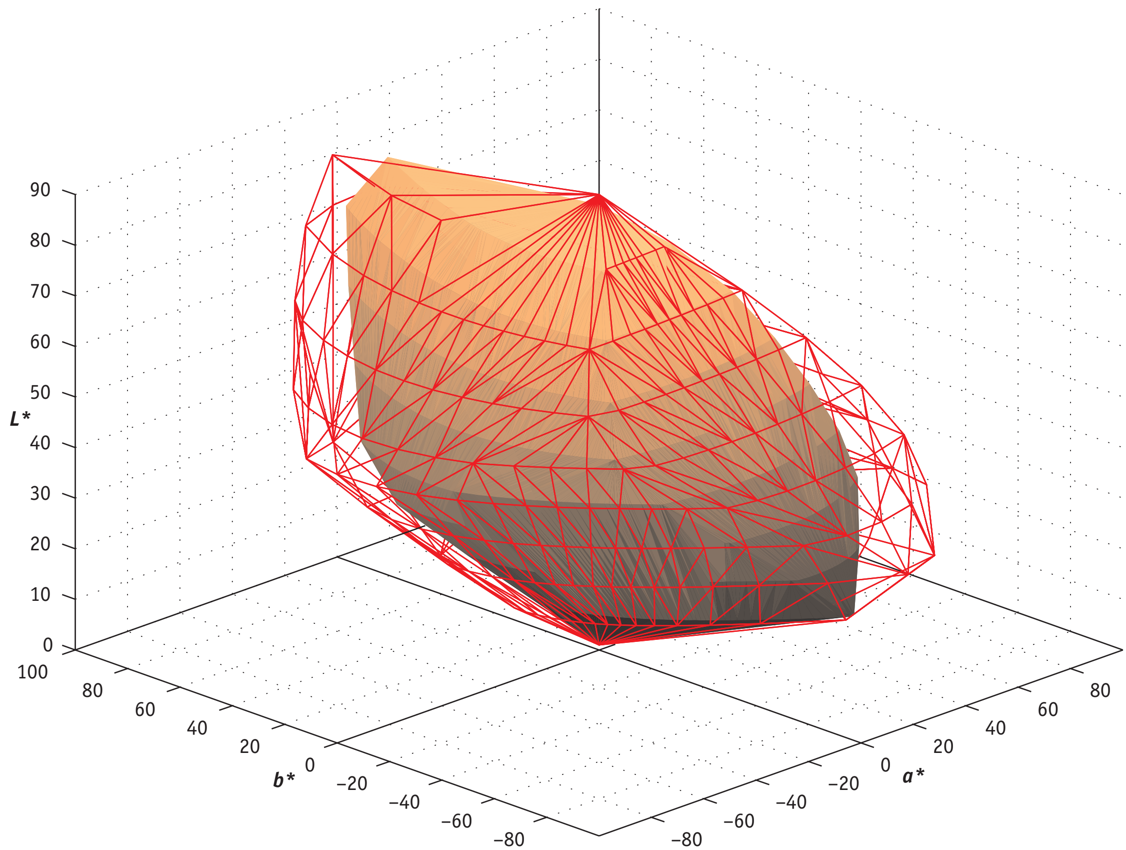


Figure 7.9c The color gamut of the dyes of Set *A* (indicated by the wire frame) is greater than that of the dyes of Set *B* (indicated by the solid).

luminance of a projected image is fairly low, and a projected image typically fills only a portion of an observer's field of view. As a consequence of these three factors, the observer's chromatic adaptation to a projected image will *not* be complete. So although a projected colorimetric neutral matches the projection illuminant in *chromaticity*, it will still appear somewhat reddish yellow.

Projection slide films therefore are designed such that the images they produce have an overall color balance that *measures* somewhat cyan-blue. This color bias, when combined with the observer's partial chromatic adaptation to the reddish-yellow projection conditions, produces images that appear to be optimally color balanced. This means, then, that they *are* optimally color balanced, given our frequently stated rule that the human observer is the ultimate arbiter of image quality.

Photographic slide films: grayscale characteristics

Figure 7.11 shows the grayscale characteristic, in terms of visual density, of a representative high-quality 35 mm photographic transparency film designed for projection. The figure also shows the grayscale of a representative reflection-print system for comparison.

Like those of the other media examined so far, the grayscale of this commercially successful film does not have a one-to-one relationship with the original scene. It is apparent from the nonlinear

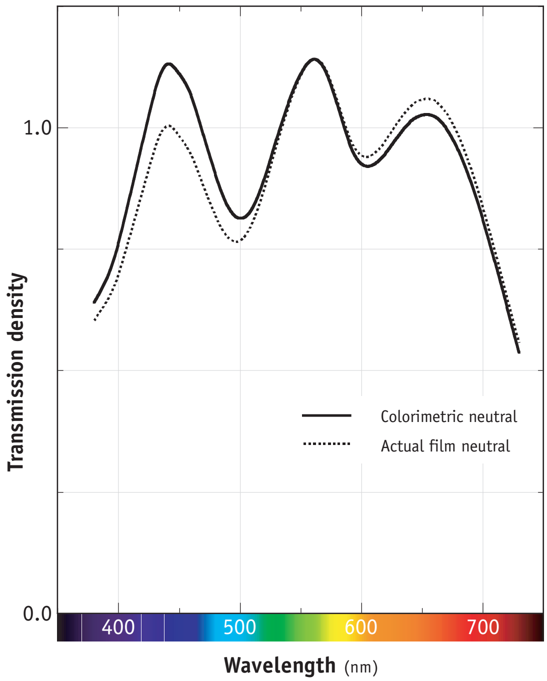


Figure 7.10 A colorimetric neutral and the actual neutral produced by a representative photographic slide film. Slides made by this film are designed to be viewed using a tungsten–halogen projection lamp.

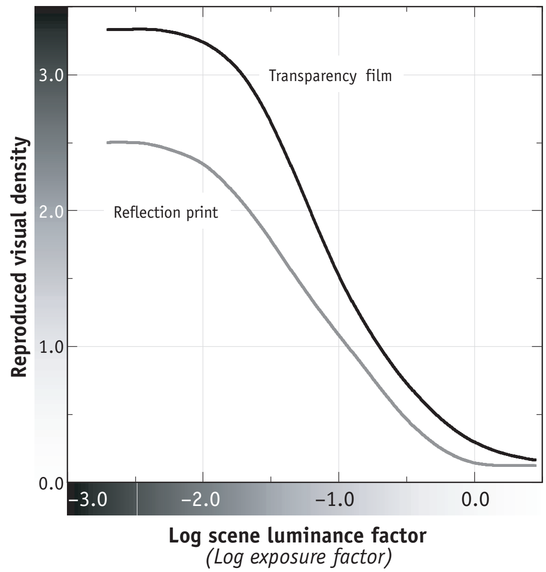


Figure 7.11 Grayscale characteristic of a high-quality photographic transparency film. The grayscale of a high-quality reflection-print system is shown for comparison.

shape—particularly from the exponential increase in visual density as the characteristic approaches higher densities—that the grayscale characteristic has been designed to compensate for viewing flare. In an otherwise darkened projection-viewing environment, viewing flare can result from stray projector light. Flare also can result from projected image light that first reflects from the projection screen to other surfaces in the room, is scattered, and then reflects back to the screen. The amount of flare light in the room, and the amount of that scattered light that reaches the image on the projection screen, will vary depending on the type of screen and the characteristics of the room. In most cases, the relative amount of flare is approximately the same as, or somewhat lower than, that encountered in video and reflection-image viewing.

Because the environment's reflection of light from the projected image back to the screen often is a major contributor of viewing flare, measurement of that flare should be made by projecting an appropriate test slide in the actual viewing environment. For example, a test-slide image might consist of a reference white patch, a dark gray patch, and a background producing an overall screen luminance level equivalent to that of typical pictorial images. The amount of viewing-environment flare then can be determined by comparing the luminance ratio of the two test-slide patches measured in the absence of flare (e.g., from densitometric values) to the luminance ratio measured by telephotometry of the projected patches from an observer's viewing position. If, for example, a 200:1 flareless luminance ratio of the test patches (white/dark) measured as only 100:1 in the projection viewing environment, the computed viewing flare would equal 0.5 % of the luminance of the white patch:

$$\frac{100}{1} = \frac{200}{1 + 200\,(pf/100)} \qquad (7.1)$$
$$pf = 0.5\,\%$$

The absolute luminance of a projected image is quite low relative to that of most original scenes. As discussed previously, this will result in the projected image being perceived as having lower luminance contrast and less colorfulness. The very high photographic gamma of the grayscale can be explained, in part, as a compensation for both of these effects. A higher grayscale photographic gamma directly increases reproduced luminance contrast. In addition, because a higher grayscale photographic gamma generally is accomplished by increasing the photographic gammas of the individual cyan-, magenta-, and yellow-dye-forming layers, reproduced color saturation also is increased.

But even if a photographic gamma increase of about 15 % (an amount comparable to that used in the electronic and reflection-print systems examined earlier) is assumed to compensate for the psychophysical effects of low absolute luminance-level viewing, and if correction for about 0.75 % viewing flare (a reasonable average) also is assumed, the high photographic gamma of the transparency grayscale still is not fully explained (Figure 7.12). In particular, the transparency grayscale is still higher in photographic gamma, and it also is higher in overall density.

Both of these characteristics of a photographic transparency grayscale are related to perceptual phenomena. The higher photographic gamma of the grayscale is required as a consequence of the viewing environment and observer *lateral-brightness adaptation*, which was described previously in Chapter 3. When a transparency image is projected in a darkened room, it essentially is surrounded by black. The black surround will cause the observer to perceive the image as having somewhat lower luminance contrast than if the image were viewed with a normal surround. Note that this reduction in apparent contrast is over and above that induced by the overall low absolute luminance levels involved. The effect of the dark surround must be counteracted by a corresponding additional increase in the photographic gamma of the transparency grayscale.

The relatively high overall density is related to another perceptual phenomenon described earlier: *general-brightness adaptation*. In this case, however, the design of the medium does not *compensate* for the perceptual effect; it takes *advantage* of it.

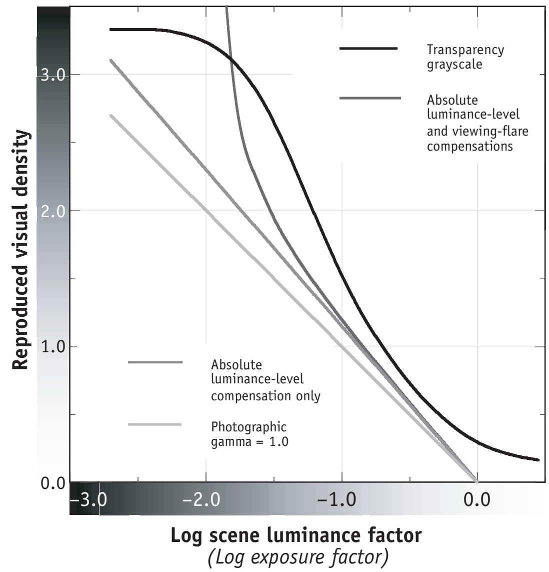


Figure 7.12 Compensations for absolute luminance level and viewing flare do not fully explain the grayscale characteristics of a photographic transparency film.

When a transparency is projected in a darkened room, there are few (if any) other visual references. As a result, the observer will adapt to a considerable extent to the absolute luminance level of the image itself. This can be demonstrated by sequentially projecting a series of slides that are identical except for their overall density. For example, the series might consist of seven slides that are identical, except that each has 0.05 greater overall density than the preceding one. Most observers, if given a brief amount of time to adapt to each image, will not detect any change in the appearances of the images from the beginning to the end of the series, even though the absolute luminance of the final projected slide is only *half* that of the first. Photographic engineers take advantage of this general-brightness adaptation effect to increase the ability of a projection slide film to record and display original-scene information.

How this is done is shown in Figure 7.13, which compares the grayscale of a photographic transparency film to that of a photographic

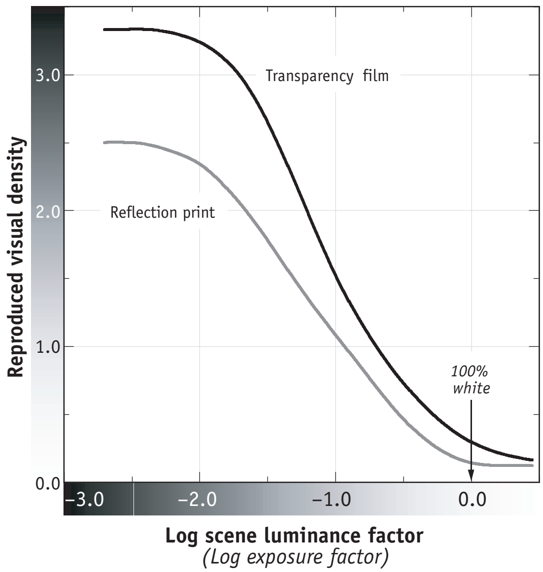


Figure 7.13 Densities of the reproductions of a perfect white on the grayscales of a photographic transparency film and a photographic reflection-print system.

reflection-print system. The figure shows that for the reflection-print system, a perfect white is reproduced very nearly at the minimum density of the print medium. Reflection images *must* reproduce whites at low densities because the environments in which such images typically are viewed contain white objects and other visual cues that essentially control the general-brightness adaptation of the observer. A reflection image with darker reproduced whites would look unacceptably dark.

However, because an observer will partially adapt to a projected image, it is acceptable to reproduce a perfect white at a relatively high density on a photographic transparency film (also shown in Figure 7.13). This makes it possible to use some of the density range of the film to reproduce highlight information corresponding to very high scene luminance-factor values. This is important, because in a typical scene, a considerable amount of important image information is present at high luminance levels, including levels that are well above the luminance of a perfect white object in the scene. Specular highlights, such as those produced by sunlight reflecting from water or polished surfaces, are one source of such information. Regions of highly illuminated diffuse highlights, such as those that will be present in certain areas of a wedding dress illuminated from above the camera angle, represent another important source of scene information above perfect white.

In addition, some areas of a scene may be more highly illuminated than the principal subject area. For example, the principal subject may be a person standing in the shade of a tree, but the scene may also include other subjects standing in direct sunlight. These sunlight-illuminated areas easily can have luminance levels well above that of a perfect white within the shaded principal subject area of the scene. Similarly, a cloudy sky may contain areas with luminances well above the luminance of a white object in the principal subject area of a scene. Fluorescent colors also may produce, at certain wavelengths, light levels greater than those reflected from a perfect white. Therefore, they too may have very high luminances. The ability of photographic transparency films to record and display such information greatly adds to the brilliance, fidelity, and almost three-dimensional appearance of images created on these media.

Photographic slide films: color characteristics

A photographic transparency film is one of the best illustrations of a point made in Chapter 2: some imaging media are complete imaging systems in themselves. Each of the basic imaging system functions—image capture, signal processing, and image formation—is performed entirely within the transparency film itself (Figure 7.14). The exact manners in which these basic functions are implemented determine the particular colorimetric characteristics of the film.

The colorimetric characteristics of the *image-capture* function of a transparency film are determined by the spectral sensitivities of that film. These sensitivities can differ somewhat from product to product, as shown in Figure 7.15. Such differences are responsible for some of the color-reproduction differences that exist among commercially available films. Figure 7.16 shows, for example, the color-reproduction properties of two photographic

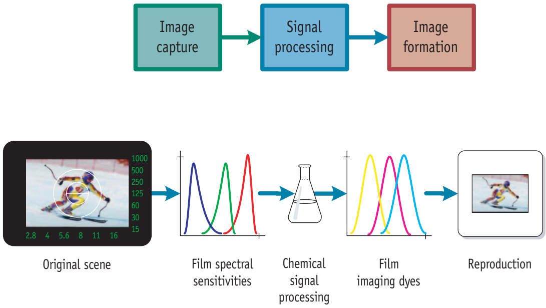


Figure 7.14 A photographic transparency film is a complete imaging system.

transparency film products that are identical in all respects except for their spectral sensitivities. The design of film spectral sensitivities is a complex process that involves the optimization of other properties, such as noise, in addition to color reproduction.

The *signal processing* function of a transparency film can be thought of as two sequential operations (although both actually occur simultaneously during chemical processing). First, the latent images, formed by exposure of the red-light, green-light, and blue-light-sensitive layers of the film, are chemically amplified. That amplification must be highly nonlinear (with respect to exposure) in order to produce the desired transparency grayscale characteristic that was shown in Figure 7.11. One way in which that visual-density grayscale can be achieved is if the density values for the constituent cyan, magenta, and yellow dye scales match those shown in Figure 7.17. The density values of that figure are expressed in terms of *equivalent neutral density* (END), which is a measure of the relative amounts of the image-forming dyes. When expressed in terms of END values, *equal* cyan, magenta, and yellow densities correspond to a colorimetric neutral for the specified viewing illuminant (see Appendix B). As the figure shows, however, throughout virtually the entire exposure range, equal *RGB* exposure-factor values produce relatively more cyan dye and less yellow dye. These unequal density relationships create the desirable cyan-blue colorimetric bias discussed earlier.

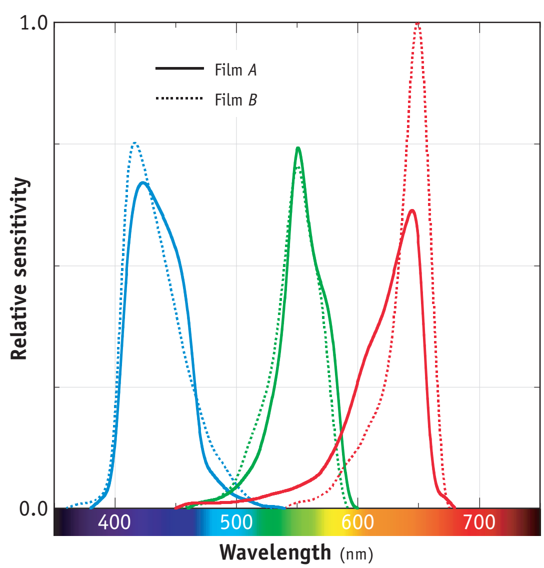


Figure 7.15 Comparison of the red, green, and blue spectral sensitivities of two photographic transparency films.

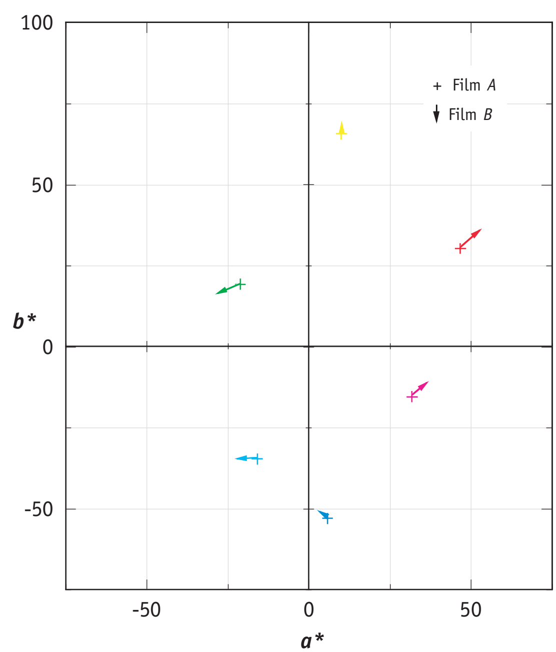


Figure 7.16 Comparison of test-target colors, as reproduced by the two films of Figure 7.15. The two films are identical except for their red, green, and blue spectral sensitivities.

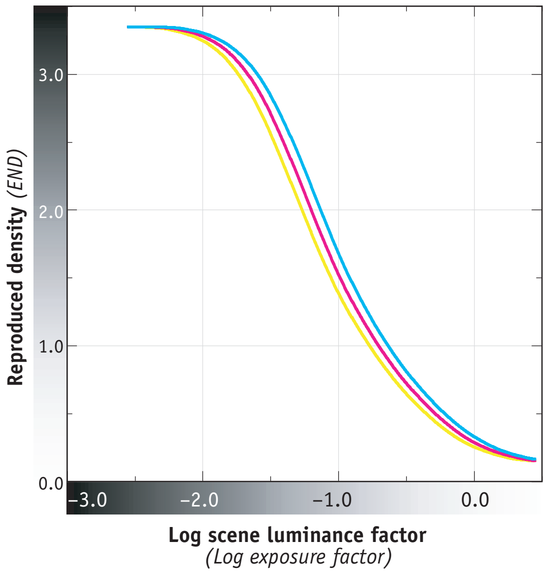


Figure 7.17 Cyan, magenta, and yellow constituent dye scales, expressed in terms of equivalent neutral density values (ENDs), for a representative grayscale of a photographic transparency film. Note the overall cyan-blue colorimetric balance.

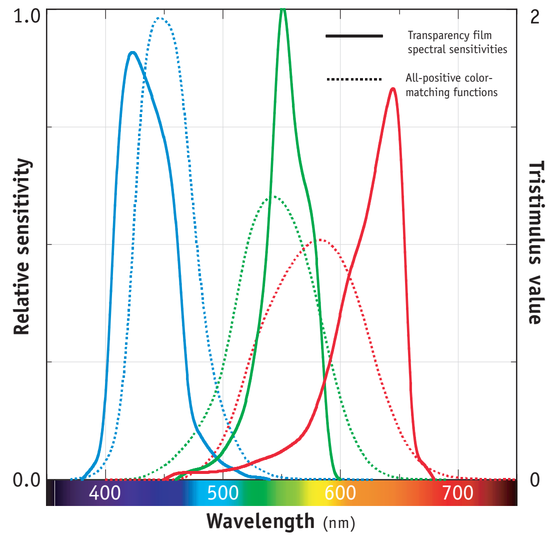


Figure 7.18 Spectral sensitivities of a representative photographic transparency film, compared to a set of all-positive color-matching functions.

The second chemical signal processing operation creates desirable interactions among the color records. The need for such interactions was explained in Chapter 6. In photographic transparency films, as well as in other silver-halide-based photographic media, *interlayer effects* are used to produce these desirable color interactions. In modern films, a variety of chemical techniques are incorporated such that image-dye formation in one film layer influences the amount of image dye that forms in one or both of the other two layers. (Refer to Appendix C for more details.)

In photographic media, chemical interactions generally behave proportionally to the *logarithm* of the exposure, rather than to the exposure itself. Logarithmic interactions are well suited for imparting application-specific alterations in color reproduction. For example, logarithmic interactions can be used to increase the overall color saturation of a film designed for advertising or to decrease the overall color saturation of a film designed for studio portraiture.

However, the transformation of captured exposure values to those that are appropriate, based on the spectral characteristics of the image-forming colorants, requires interactions that are *linear* with exposure, as was explained for the electronic color-imaging system described in Chapter 5. Chemical interactions that behave linearly with respect to exposure are extremely difficult to achieve in photographic media. Film spectral sensitivities therefore are designed such that the *need* for these interactions is minimized. This can be accomplished by the use of spectral sensitivities that differ somewhat from visual color-matching functions. For example, Figure 7.18 compares the sensitivities of a representative photographic transparency film to a set of color-matching functions. The film sensitivities are spectrally narrower and more separated than this or any other set of all-positive color-matching functions.

These departures from color-matching functions must be designed very carefully in order to optimize the overall color reproduction of the film. Particular attention has to be paid to the reproduction of important memory colors, such as human skin tones, foliage, and sky. The reproduction of metameric pairs also is an important consideration in the design of film spectral sensitivities. Important metameric pairs should produce matched reproductions.

Photographic motion picture films

Photographic motion pictures can be made using positive-working films that are similar to those used for photographic transparencies. In practice, however, major theatrical motion picture releases instead are produced using combinations of negative-working photographic media for image capture and display. There are many advantages to this approach. In particular it facilitates editing and the addition of special effects, and it makes practical the production of large numbers of projection prints.

In an entirely optical system, original scenes are captured on a camera-speed photographic negative film. That negative then can be optically printed directly onto a negative-working motion picture print film to produce a positive image for projection. Typically, however, the original camera negative first is duplicated by printing it onto a negative intermediary film to produce what is called a *master positive*. The master positive is then itself printed onto the same type of intermediary negative film to produce what is called a *master negative*. Finally, the master negative is used to print onto motion picture print film to produce multiple prints for projection in theaters (Figure 7.19). Because the master negative is a virtual duplicate of the original camera negative, the end result of this production process is projection prints having colorimetric characteristics substantially the same as those of a print made directly from the original camera negative.

In practice, the production process also may involve other operations. For example, multiple master positives and master negatives may be produced, films may be scanned and digitized for editing and the addition of special effects, and numerous other steps may be included in the production of the final theatrical prints. Such processes, work flows, and other issues related to motion picture production will be described in detail in Chapter 21. For now, the emphasis is on the colorimetric properties of the projected images.

The grayscale characteristic of a motion picture print from a representative production system is shown in Figure 7.20. Note that this overall system

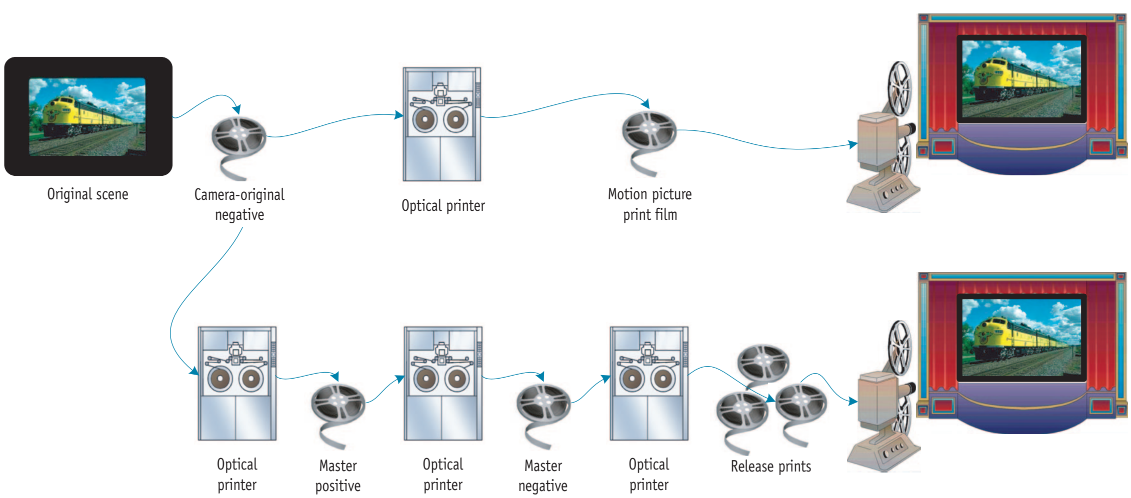


Figure 7.19 An optical motion picture color-imaging system. Camera negatives can be optically printed directly onto a motion picture print film to produce a positive image. More commonly, the original negative is first duplicated to produce one or more duplicate negatives from which large numbers of theater projection prints are made.

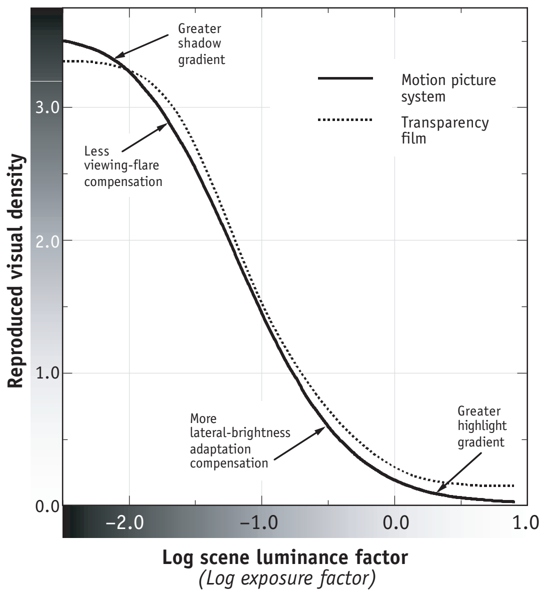


Figure 7.20 Representative grayscales, expressed in terms of visual density values, for an optical photographic motion picture system and a photographic transparency film.

grayscale has much in common with the grayscale of a photographic transparency film. This is not surprising since the respective systems also have much in common. Nevertheless, there are some differences between the two grayscale characteristics. The differences illustrated in Figure 7.20 can be attributed primarily to the following distinctions between the motion picture and transparency slide-film systems:

- The viewing environment of a motion picture theater typically has less viewing flare than that of most slide-film viewing environments, in part because theater designs attempt to minimize the reflection of image light back to the screen. Accordingly, less viewing-flare compensation is included in the motion picture grayscale characteristic, as can be seen in the higher-density regions of the two characteristics.
- A motion picture image surround typically is darker than that of most slide-film viewing environments, so somewhat more compensation for observer lateral-brightness adaptation is included in the motion picture grayscale characteristic. The required higher overall luminance contrast of the motion picture grayscale is most apparent in the mid-scale and lower mid-scale regions of the curves, where opposing contrast differences due to the different flare correction levels of the two systems are less dominant.
- The original-scene luminance dynamic range that can be recorded by a negative-working camera film is greater than that recordable by a positive-working transparency film. The overall motion picture system grayscale characteristic is designed to take advantage of that greater image-capture capability. This is apparent in the greater density gradients of the extreme shadow and highlight regions of the motion picture system grayscale.
- The luminance dynamic range (visual density range) that can be produced on a negative-working motion picture print film is somewhat greater than that achievable on a positive-working transparency film. The overall motion picture system grayscale characteristic is designed to take full advantage of that greater image-display capability. This difference again is apparent in the greater density gradients of the extreme shadow and highlight regions of the motion picture system grayscale.

In addition, motion picture theater projection lamps operate at color temperatures higher those of slide-film projectors. As a consequence, observer chromatic adaptation to the chromaticity of the motion picture projector lamp is somewhat more complete. Therefore, the color balance of a motion picture print is designed to be more nearly colorimetrically neutral than that of a slide film, as illustrated by Figures 7.21a and 7.21b.

Electronic motion picture projection

The topic of digital electronic motion picture systems will be discussed in detail later, in Chapter 21. For the time being, it is sufficient to recognize that digital cinema projectors can be used in motion picture theaters to display images originally captured on photographic films, images originally captured

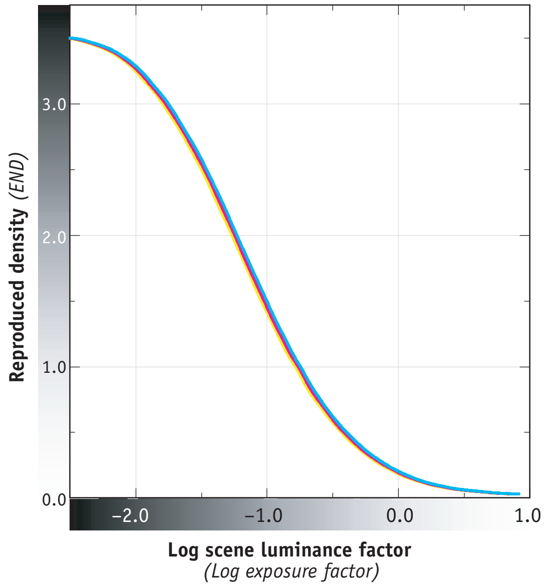


Figure 7.21a Cyan, magenta, and yellow constituent dye scales, expressed in terms of equivalent neutral density values (ENDs), for a representative grayscale of an optical photographic motion picture system.

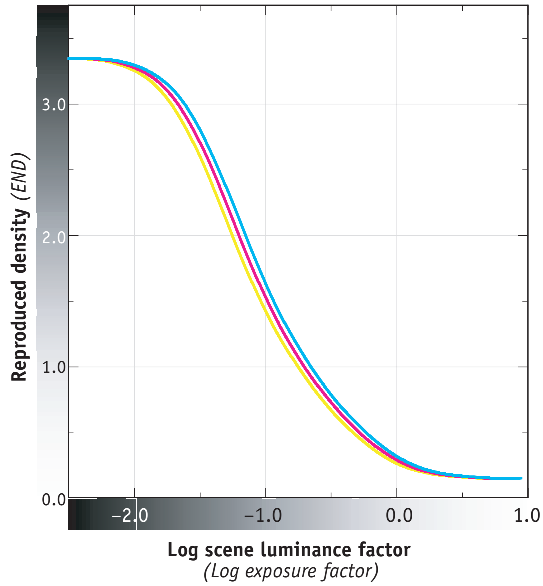


Figure 7.21b Cyan, magenta, and yellow constituent dye scales, expressed in terms of equivalent neutral density values (ENDs), for a representative grayscale of a photographic transparency film.

by electronic cameras, images generated using computer graphics, or any combination of the three.

Moreover, it is most important to understand that the fundamental colorimetric characteristics required for projected motion picture images ultimately is not determined by the methodology used for image capture or origination, nor is it determined by the particular technology used for final image display. Instead, it is determined by the characteristics of the theater *environment* in which the images are displayed and viewed. It will be shown later that a generalization of this important concept of viewing-environment-dependent colorimetry can provide the basis for a universal system of color management.

Other modes of dark-surround image viewing

The two systems discussed to this point—slide film and motion picture—are both intended for front projection in a darkened environment, but other modes of image display also may correspond to dark-surround viewing. These modes can include various technologies for rear-projection, rear-illumination of photographic transparencies, and self-luminous displays.

It is important to recognize that a given display technology or medium may not necessarily be associated with one particular type of viewing environment. For example, a large-format photographic transparency could be back-illuminated such that its average luminance level is greater than that of the rest of the environment, or it could be illuminated such that its average luminance level is similar to that of the environment. In the first case, the viewing surround would be considered dark; thus the optimum image colorimetry essentially would correspond to that of a photographic slide film. In the second case, however, the viewing surround would be considered normal rather than dark. Therefore different image

Figure 7.22 An identical image presented with two different surrounds. The image appears to change in lightness and in luminance contrast, which again demonstrates that a specification of an image's colorimetry alone is not sufficient to describe its appearance.

colorimetry, corresponding somewhat more closely to that of a reflection print, would be preferred.

Most reflection prints are displayed in normal-surround viewing environments. However, a reflection print could be displayed in a dark-surround environment by placing it on a dark background and illuminating it with a spotlight that illuminates only the area occupied by the print. In that case, the print colorimetry ideally should correspond to the characteristics of a photographic slide-film or motion picture system.

Television display environments generally are considered to be "dim surround," i.e., the display ordinarily is somewhat brighter than the adjacent illuminated areas of the viewing environment. Adjusting the ambient room illumination obviously can change the environment from dim-surround to normal-surround to dark-surround conditions, and this may require corresponding adjustments to image colorimetry. In home video systems, this can be accomplished using the brightness, contrast, and chroma adjustments incorporated in the display device.

Images displayed on computer monitors generally are less sensitive to changes in the room environment because the observer's field of view typically is dominated by the display itself due to the relatively short viewing distances involved. However, the image surround can be altered significantly if the image itself does not fill the entire display and the areas surrounding the image are changed in relative luminance. Figure 7.22 illustrates how the appearance of a colorimetrically identical image changes when the surrounding areas of the image display are white and black. Note that the image appears to change in both lightness and luminance contrast. This again demonstrates that a specification of image colorimetry alone is not sufficient to describe color appearance.

Color-encoding considerations

This analysis of the colorimetric characteristics of projected and other forms of dark-surround viewed images has uncovered several important factors related to color encoding. Of particular importance is the fact that when images are viewed in a dark-surround environment, the relationship between their colorimetric measurements and their visual appearance is not straightforward. For example, standard colorimetric measurements would indicate that photographic transparencies intended for projection are too dark, too high in luminance contrast, and cyan-blue in overall color balance compared to an original scene or a reflection image—yet they do not *appear* that way at all. This discrepancy between colorimetric measurement and color appearance is a very significant complication that will be addressed later, when several alternative methods for encoding color-image information are discussed.

Another factor related to color encoding is that the colorimetric characteristics of images on media and devices that may be used for input to digital color-imaging systems are produced by a complex set of relationships involving the characteristics of spectral sensitivities, grayscale signal processing, color signal processing, and image-forming colorants. One consequence of this complexity is that various types of input, and even different products of the same basic type, tend to have their own distinctive appearances. Sometimes the color-reproduction characteristics that contribute to a particular appearance are created deliberately, and sometimes they are a result of various design compromises. Regardless of their origin, the color-reproduction differences among various inputs are significant. For some applications, these differences are desirable and should be retained in the image color encoding. However, for other applications such as image compositing, such colorimetric and color-appearance differences among the inputs can be problematic. In any case, the differences among inputs are real; thus they must be taken into account in any digital color-encoding method that supports multiple types of inputs.

There are additional considerations that apply specifically to photographic transparency films and photographic projection print films. First, they have very large color gamuts, in part because their image dyes are formed on transmissive, rather than reflective, supports. Because they are designed for dark-surround viewing, by necessity they also must have extensive luminance dynamic ranges. Successful color encoding of images on these media therefore must be capable of numerically representing large color gamuts and luminance ranges. In addition, the image-forming dyes used in some transparency media are very viewing-illuminant sensitive. Accordingly, these media are designed specifically for one particular viewing illuminant. This is an important consideration in scanning images on such media for input and color encoding in that scanned colorimetric values will be strongly influenced by the particular spectral properties of the scanner illuminant.

Summary of key issues

- Like reflection images, images on transparency media are objects that produce color stimuli only when illuminated.
- Some transparency media are quite sensitive to changes in viewing-illuminant spectral power distribution and thus may be designed for use with one specific viewing illuminant.
- Different photographic transparency films and other media have different characteristic appearances that result from their particular capabilities and limitations. Other differences are created deliberately, based on the specific applications for which the media are intended.
- Transparency media often have very large color gamuts, in part because their image dyes are formed on transmissive supports. Successful color encoding of these media therefore must be capable of numerically representing large color gamuts.
- When an image is viewed in an otherwise darkened environment, the observer's state of adaptation is influenced by the image itself, and the appearance of that image is influenced by its dark surround. Under such conditions, the relationship between standard colorimetric measurements and visual appearance is not straightforward.
- Projected and other forms of dark-surround viewed images must depart significantly from a one-to-one colorimetric relationship with the original scene in order to compensate appropriately for viewing flare and for observer general-brightness adaptation, lateral-brightness adaptation, and incomplete chromatic adaptation.
- Images produced for projection and other forms of dark-surround viewing must have extensive luminance dynamic ranges. Successful color encoding of such images therefore must be capable of numerically representing dynamic ranges considerably larger than those required for other types of images.

8

Photographic Negatives

Although they generally are designed for a different purpose—to be optically printed onto photographic papers and print films—photographic negative films have inherent properties that make them well suited for input to digital color-imaging systems. In particular, they can record (and make available to a scanner) color information from an extremely wide range of exposures. On many photographic negative films, the dynamic range of recorded exposures can easily exceed a ratio of 10 000:1 (a 4.0 log exposure range).

Dynamic range is one reason why color negative films remain a dominant image-capture medium for professional motion pictures. Their continued use in that application is one reason why understanding their imaging properties is still relevant. There are other reasons as well. For one, there is a legacy of negative images that will need to be digitized and stored, and doing this properly requires an understanding of the unique properties of these media. In addition, negative films represent a class of images that can be considered *intermediary*, i.e., their image values do not directly represent either the colorimetry of the photographed scene or the colorimetry of the image that ultimately will be produced for viewing. This is an important concept that, as will be shown later, applies to other forms of images and must be considered in the design of any comprehensive imaging system or color-encoding method.

Negative films are low in photographic gamma. So despite their extensive exposure dynamic range, they produce relatively low optical densities. For example, a 100:1 ratio of exposures (a 2.0 log exposure range) would result in a density range of approximately 3.0 on a representative photographic transparency film, but that same ratio of exposures would result in a density range of only about 1.1 on a representative photographic negative film (Figure 8.1). This is a desirable characteristic for scanning, particularly in cases where the dynamic range of the scanner's sensor is somewhat limited and/or there are scanner signal-to-noise problems. In certain types of scanners, it is quite difficult to derive meaningful information from scanned areas of very high densities.

Color photographic negatives do have one important disadvantage: they are not easily "human readable." Their low contrast and overall orange color cast make them difficult to evaluate visually. Even more troublesome is that they produce images that are "backwards." They reproduce whites as blacks, blacks as whites, reds as cyans, greens as magentas, blues as yellows, and so on. It requires a great deal of experience and skill to judge negative images by direct viewing.

The lack of human readability of color negatives is an important concern in applications where large numbers of images must be evaluated for selection. In the production of catalogs, for example, perhaps a hundred images might be examined for every one that is actually chosen for use. In some national

Digital Color Management: Encoding Solutions, Second Edition Edward J. Giorgianni and Thomas E. Madden
 ISBN: 978-0-470-51244-9

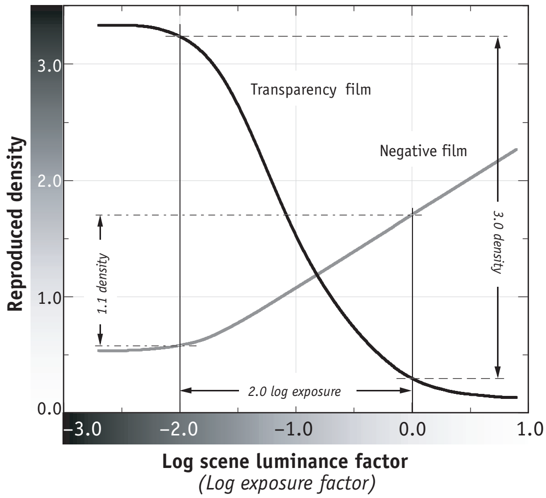


Figure 8.1 Comparison of log exposure ranges and corresponding density ranges of negative and transparency photographic films.

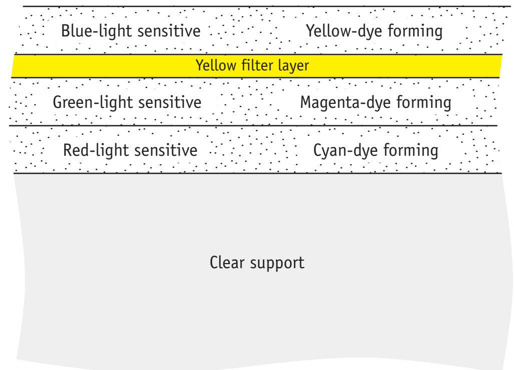


Figure 8.2 Simplified cross-section of a color-negative photographic film (prior to chemical processing).

magazines, the ratio of images examined to those used can be more than a *thousand* to one.

At one time, the scanning of photographic media was slow and expensive, making it impractical to scan large numbers of images for evaluation and selection. Image selection therefore had to occur *prior* to scanning. That is one of the principal reasons why (positive) transparency films, which are problematic to scan but can be directly viewed, generally were favored over negative films for input in the majority of film-based digital imaging applications. The current situation is quite different. High-quality scanning is fast and inexpensive, and many digital imaging systems provide accurate softcopy previews of both positive and negative scanned images. This essentially eliminates the need to print negatives simply for the purpose of evaluation.

General characteristics

Figure 8.2 shows a simplified cross-section of a photographic negative film. The basic structure is identical to that of a photographic transparency film: a blue-light-sensitive layer, yellow filter layer, green-light-sensitive layer, and red-light-sensitive layer coated on a clear support. When the exposed film is chemically processed, yellow, magenta, and cyan dyes are formed in the blue-light-sensitive, green-light-sensitive, and red-light-sensitive layers, respectively. Unlike a photographic transparency film, however, a photographic negative film produces its *maximum* amount of image dye from a *maximum* exposure and its *minimum* amount of dye from a *minimum* exposure.

If the structures are the same, why does a photographic transparency film produce a positive image and a photographic negative film produce a negative image? It is a common misconception that this happens because photographic negative films somehow “see” the world backwards. But that is not the case.

Photographic negative and transparency films actually “see” the world in much the same way. In other words, their sets of red, green, and blue spectral sensitivities are not fundamentally different. The sensitivities may, in fact, be virtually identical for a given pair of negative and transparency films. In terms of basic imaging system functions, then, these two media are not fundamentally different in their *image-capture* characteristics. Both types of media form cyan, magenta, and yellow dyes, so they also are not fundamentally different in their *image-formation* characteristics.

That leaves just one imaging system function—*signal processing*—that must be responsible for making a photographic film produce either a negative or a positive image. That can easily be demonstrated with a simple experiment. If a transparency film is exposed and then processed in a chemical process intended for negative films, its images will come out *negative*, not positive. Likewise, if a negative film is exposed and then processed in a transparency film chemical process, its images will come out *positive*, not negative.

Neutral characteristics

Figure 8.3 shows the spectral transmission densities of the cyan, magenta, and yellow image-forming dyes for a representative photographic negative film. Note that the spectral characteristics of these dyes, particularly the cyan dye, are somewhat different from the spectral characteristics of the dyes of a representative photographic transparency film. That negative and transparency media form somewhat different dyes might be expected, since transparency films are designed to be viewed by humans and negative films are not. This fact raises an interesting issue.

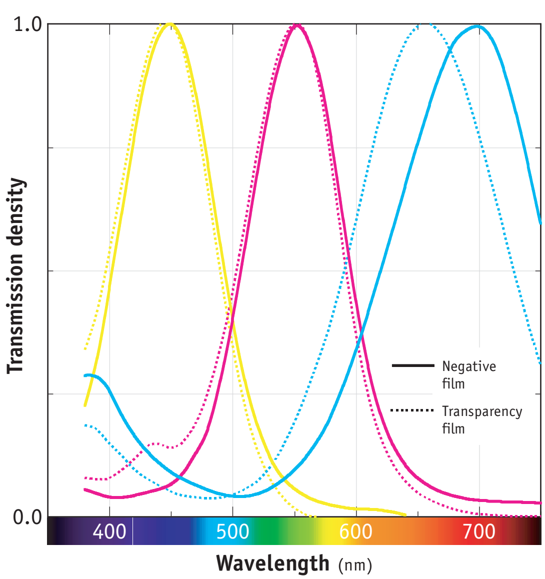


Figure 8.3 Comparison of the cyan, magenta, and yellow image-forming dyes of a representative photographic color-negative film (solid lines) and a transparency film (dotted lines).

In previous system examinations, standard colorimetric measurements were used to quantify the trichromatic characteristics of color stimuli produced by the output media. It was logical to make such measurements because the media involved formed positive images intended for direct viewing. But since negatives are not meant to be viewed directly, it would not seem logical to also measure them using methods based on the responses of a human observer.

If standard CIE colorimetry is ruled out, how else can images on negative films be measured? Is there a different type of "standard observer" that would be more appropriate for negative images? Answering these questions requires a closer look at how photographic negative films are designed and how they are used in practice.

In typical applications, photographic negatives are optically printed onto a second negative-working photographic medium (Figure 8.4). The second medium might be photographic paper, in which case the final image is a reflection print. Negatives also can be printed onto special clear-support films, such as those used to make motion picture projection prints from motion picture negatives. In either case, the resulting print is a directly viewable positive image.

What is important to appreciate is that each photographic negative film is designed to be optically printed onto one or more *specific* print films or papers, using *specific* printer light sources. That fact provides the key to making meaningful measurements of color-negative images.

Optical printing is an image-capture process that is quite similar to several others discussed earlier. As shown in Figure 8.5, there is an object (the negative image) that is illuminated (by the printer light source) and "viewed" by an "observer" (in this case, a print medium). The print medium for which the negative film is intended, then, should be considered the "standard observer" for that film, and measurements should be made *according to the particular red, green, and blue spectral responsivities of the intended print medium.* In other words, measurements should be made based on what the intended print medium will "see" and *capture* when it "looks" at the illuminated negative in the printer.

What the print material will capture, i.e., the red, green, and blue *exposures* it will form, can be

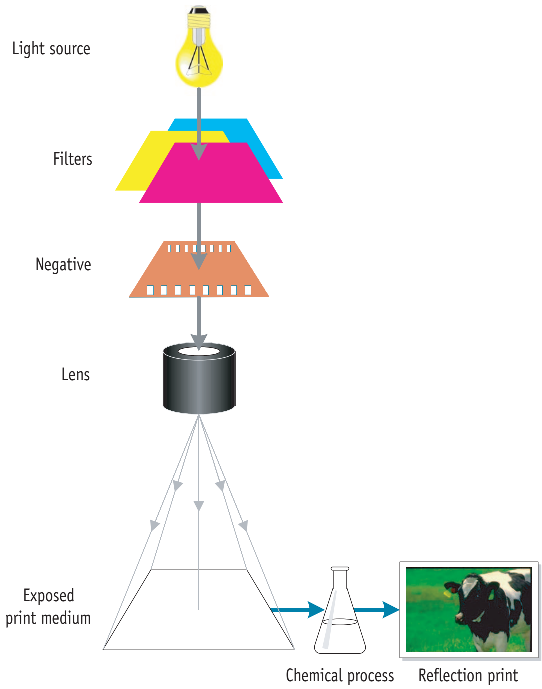


Figure 8.4 Optical printing of a photographic negative onto a photographic print medium. The negative is illuminated by a printer light source. Color filters are used to control the overall density and color balance of the print by altering the spectral power of the printer light. Filtered light passes through the negative and is focused onto a photographic print medium. Finally, the exposed print medium is chemically processed to form a directly viewable image.

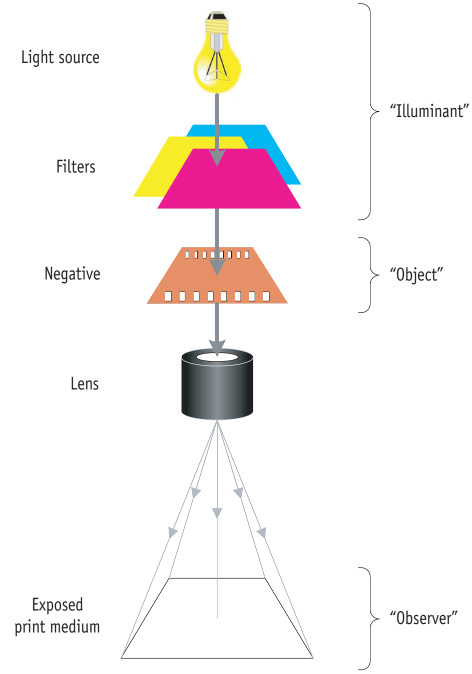


Figure 8.5 The optical printing of a photographic negative film onto a photographic print medium is an image-capture process involving an object, an illuminant, and an "observer."

computed using the following equations:

$$\begin{aligned} R_{exp} &= k_r \sum_{\lambda} S(\lambda) T(\lambda) r(\lambda) \\ G_{exp} &= k_g \sum_{\lambda} S(\lambda) T(\lambda) g(\lambda) \\ B_{exp} &= k_b \sum_{\lambda} S(\lambda) T(\lambda) b(\lambda) \end{aligned} \tag{8.1}$$

where R_{exp}, G_{exp}, and B_{exp} are the print-medium red, green, and blue exposure-factor values; $S(\lambda)$ is the spectral power distribution of the printer light source; $T(\lambda)$ is the spectral transmittance of the negative "object"; $r(\lambda)$, $g(\lambda)$, and $b(\lambda)$ are the red, green, and blue spectral sensitivities of the print medium; and k_r, k_g, and k_b are normalizing factors. These factors usually are determined such that R_{exp}, G_{exp}, and $B_{exp} = 1.0$ for a theoretical 100 % transmission (zero optical density) negative.

The negative logarithms of the red, green, and blue exposure-factor values are called *printing densities*:

$$\begin{aligned} PD_r &= -\log_{10}(R_{exp}) \\ PD_g &= -\log_{10}(G_{exp}) \\ PD_b &= -\log_{10}(B_{exp}) \end{aligned} \tag{8.2}$$

where PD_r, PD_g, and PD_b are the red, green, and blue printing-density values, and R_{exp}, G_{exp}, and B_{exp} are the red, green, and blue exposure-factor values that were calculated from Equations (8.1).

The neutral and color characteristics of all color-negative films are *designed* in terms of computed printing densities; therefore, printing densities are the most meaningful densitometric measurements that can be made from color-negative films. One way to conceptualize the measurement of printing densities is to imagine a densitometer having a light source that is spectrally identical to the printer light source, and red, green, and blue light sensors having spectral sensitivities identical to those of the print material (Figure 8.6a). A densitometer meeting those criteria would *directly* measure printing densities.

In practice, such specialized densitometers generally are not used. Instead, printing-density values are computed from measured spectral transmission data, as in Equations (8.1) and (8.2). Printing-density values also can be determined from appropriate mathematical transformations of density values measured with an ISO Status M densitometer, as shown in Figure 8.6b.

Just as it is possible to have metameric neutrals for human observers, it is possible to have metameric neutrals as "seen" by print materials. If color areas on two negative films have *the same printing densities* (determined for a particular print material and printer light source), both areas will produce *the same color* on the final print when the negatives are printed identically, using that particular printer light source and print material. This will be true even if the spectral density characteristics of the image-forming dyes of the two films, and thus the spectral characteristics of all color areas formed by the two films, are different. For example, Figure 8.7 shows the different spectral characteristics of metameric color areas from two different negative films.

Printing-density metamerism makes it possible for manufacturers to design negative films with different sets of CMY image-forming dyes for use with the same print medium. Figure 8.8, for example, shows the dye sets used in two different color-negative films. Despite these differences in their dye sets, both films will print satisfactorily onto the same photographic paper medium because each was designed in terms of printing-density values based on that medium and a specified printer light source.

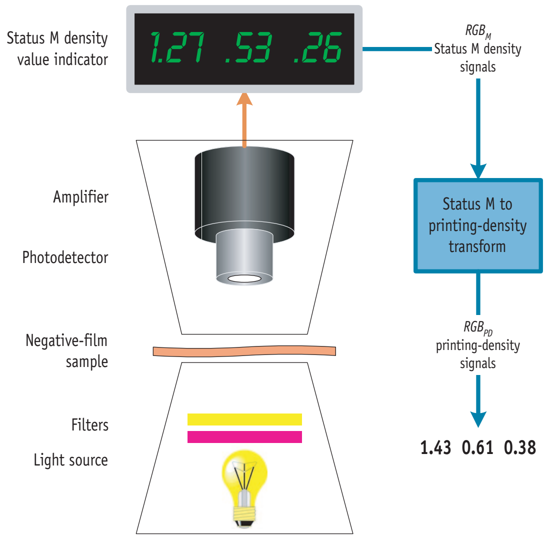


Figure 8.6b Printing-density values also can be measured using an ISO Status M densitometer and an appropriate mathematical transformation. A different transformation is required for each different combination of negative film, printer, and print material.

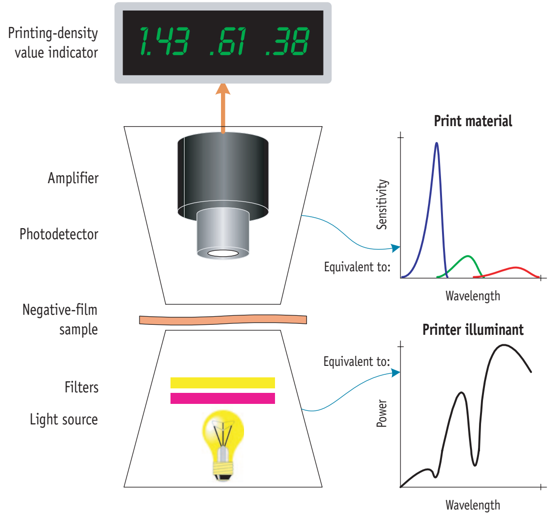


Figure 8.6a This specialized densitometer would directly measure printing densities for the specified printer and print material.

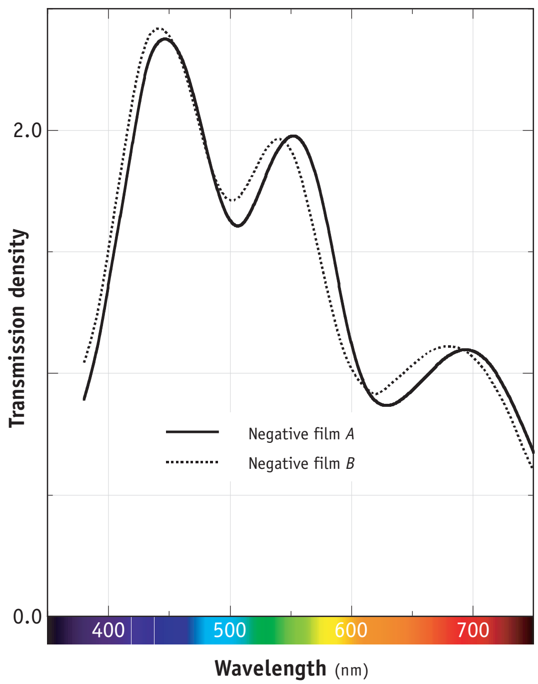


Figure 8.7 Printing-density metamerism. The spectral density characteristics of two areas on two different color-negative films are quite different. However, they have the same red, green, and blue printing densities for a particular print medium and printer.

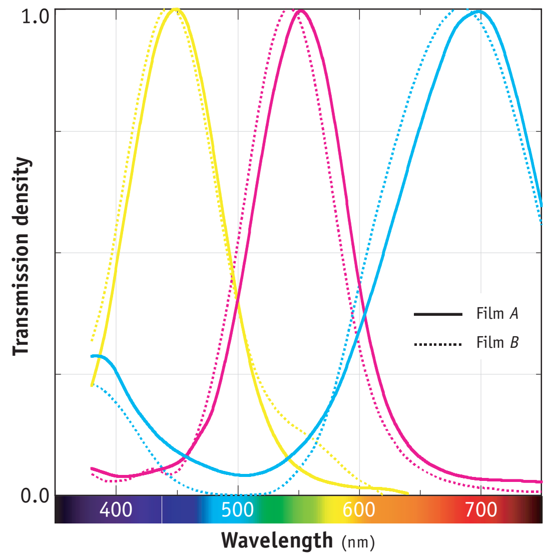


Figure 8.8 Image-forming dye sets used in two different color-negative films.

Grayscale characteristics

Figure 8.9a shows, in terms of printing densities, the grayscale characteristic for a representative photographic negative film. The curves show the basic negative-working characteristic of the film: greater exposure levels result in greater density produced in the processed film. The curves also show the extensive exposure dynamic range and low photographic gamma of the film. The somewhat higher overall densities of the blue and green printing-density curves are indicative of the overall orange color cast of negative films. That color cast, which is part of the color-correction signal processing built into the film, is compensated for in the printing process. (Refer to Appendix C for a more detailed explanation of the purpose of this orange cast.)

Note that the red, green, and blue printing-density curves of Figure 8.9a are perfectly parallel to one another. They differ only in terms of their overall printing-density value. This means that

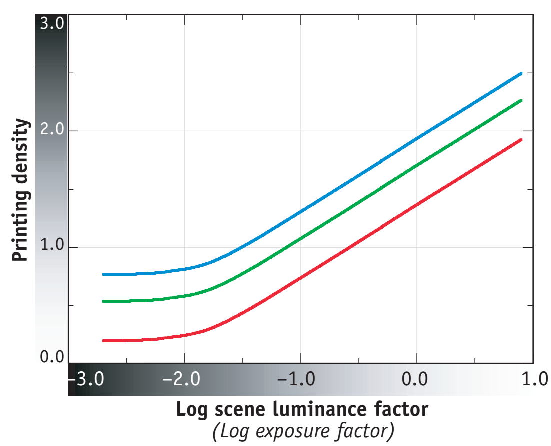


Figure 8.9a Grayscale characteristic, plotted in terms of printing densities, for a representative photographic negative film. Note that the log exposure-factor range is 4.00 (a 10 000:1 ratio of linear exposure-factor values).

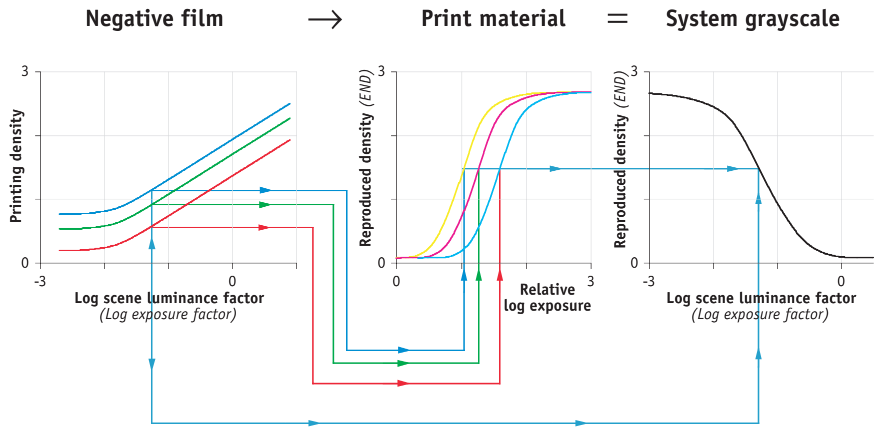


Figure 8.9b The grayscale printing-density curves for the film of Figure 8.9a are perfectly parallel to each other. This means that when the negative is printed properly, a printed grayscale will be perfectly neutral throughout the scale.

when the negative is printed properly onto an appropriate print medium, a printed grayscale will be perfectly neutral from one end of the scale to the other (Figure 8.9b).

The use of metrics other than printing density for the measurement of photographic negatives can be misleading. For example, Figure 8.10a compares red,

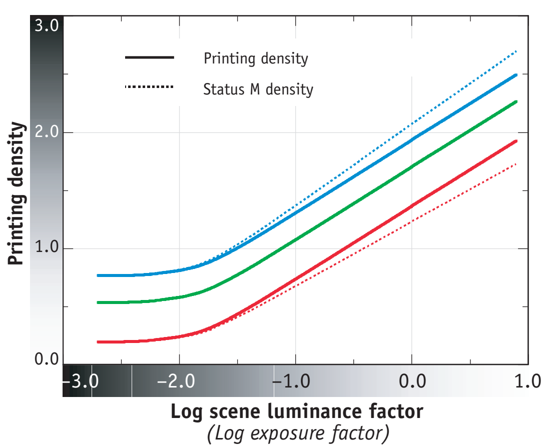


Figure 8.10a Grayscale characteristic for a color-negative film, measured in terms of printing-density values and in terms of Status M density values.

green, and blue printing-density curves to Status M density curves for the same negative-film grayscale. In terms of Status M density values, the grayscale curves are no longer parallel. The Status M measurements therefore incorrectly imply that when the negative is printed properly onto an appropriate print medium, a printed grayscale will *not* be perfectly neutral throughout the scale. As shown in Figure 8.10b, the Status M measurements incorrectly imply that the print grayscale will be yellow-red at higher densities and cyan-blue at lower densities. For proper interpretation, then, Status M density values should be transformed to printing-density values.

Like photographic transparency films, negative films generally are color balanced for use with particular reference scene illuminants. However, the optical printing process provides an opportunity to adjust the overall density and color balance of the final print made from a negative. The optical printing step thus allows greater flexibility in the design and use of the film.

For example, if a negative photographic film is overexposed, the overall density of the resulting processed negative will be too high. But when that negative is printed, more printer light can be used (by increasing the intensity of that light and/or by increasing the printing time). This will result in a print

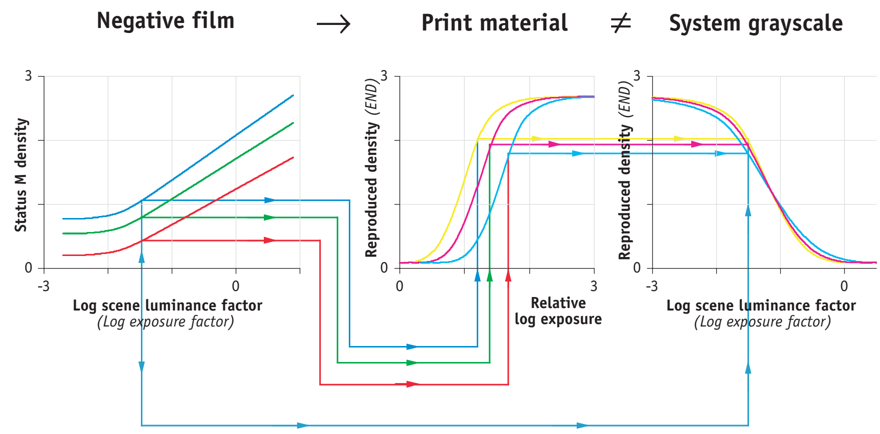


Figure 8.10b The Status M density grayscale curves shown in Figure 8.10a are not parallel to each other. This incorrectly implies that if the negative were printed, the printed grayscale would not be perfectly neutral throughout the scale.

that is properly balanced in overall density. Similarly, if a negative is exposed using a light source that is too blue, the overall blue density of that negative (the amount of yellow dye formed) will be too high. But if more blue light is used in printing, a properly color-balanced print again will be produced.

Because color-balance corrections can be applied during the output printing process, the color balance of a negative-film grayscale is far less critical than that of a film—such as a photographic slide film—that is not printed, but is instead viewed directly. Since their color balance is not particularly critical, negative films can be designed with grayscales that are color balanced such that good results are produced when the films are used to capture images in any of a variety of scene-illumination conditions. For example, a photographic negative film that is labeled "daylight balanced" actually may be designed with extra sensitivity to blue light (Figure 8.11a).

This additional blue sensitivity will result in too much blue-light exposure when the film is used to photograph daylight-illuminated scenes. But it also helps to prevent blue-light underexposure when the film is used to photograph tungsten-illuminated scenes. This is a sensible compromise. While a negative film can record information that is considerably overexposed, and an overexposed negative can be compensated for in printing, underexposure results in an unrecoverable loss of scene information (Figure 8.11b).

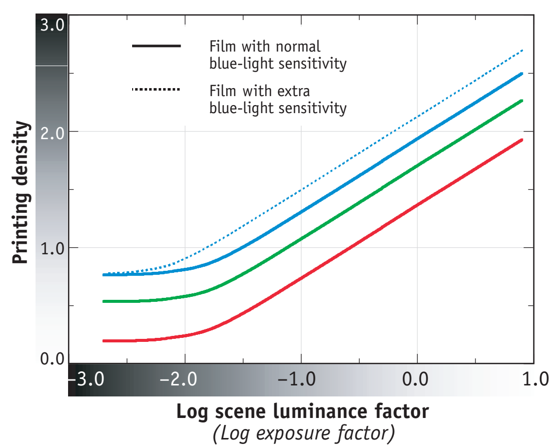


Figure 8.11a A "daylight-balanced" negative film may be designed with extra sensitivity to blue light. The extra blue sensitivity helps prevent blue-light underexposure when the film is used to photograph tungsten-illuminated scenes.

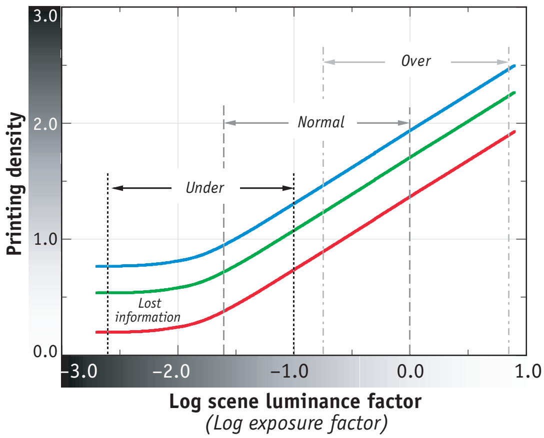


Figure 8.11b A negative film can record information that is considerably overexposed, but underexposure results in an unrecoverable loss of original-scene information.

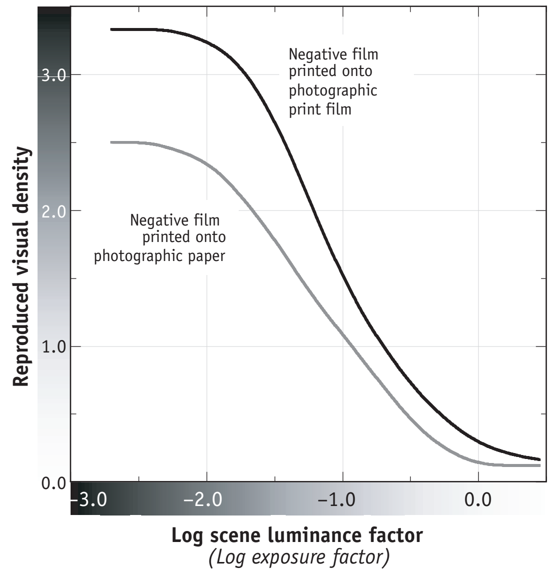


Figure 8.12 System grayscales that result from printing a photographic negative onto photographic paper and a photographic print film.

The grayscale characteristics shown thus far are representative of those of many color-negative films. However, the characteristics of individual products will differ somewhat, based primarily on their intended application. For example, negative films intended for studio portraiture generally are lower in photographic gamma. Films intended for amateur systems tend to be higher in photographic gamma, in part to help compensate for the higher flare and lower sharpness of lower-cost cameras. In addition, it has been found that many amateurs tend to prefer higher-contrast, "snappier" looking prints.

When a typical color photographic negative is printed onto a typical photographic reflection-print material, the resulting overall system grayscale characteristic, relating the original scene to the final print, will be similar to that described in Chapter 5 and shown again here in Figure 8.12. When a typical color negative is printed onto a typical photographic transmission print material, the resulting system grayscale characteristic will be similar to that described in Chapter 6 and also shown here in Figure 8.12.

The basic relationships of the grayscale characteristics for a negative, a print medium, and an overall system were shown previously in Figure 8.9b. Achieving system grayscales like those of Figure 8.12 requires proper design of both the negative and print medium. The photographic gamma of the system grayscale characteristic is simply the product of the photographic gammas of the negative and the print medium. However, because negative grayscales are straight-line characteristics when plotted in terms of density versus log exposure factor, the desirable nonlinear characteristics for the overall system grayscale, also plotted in those terms, must be designed into the print medium rather than the negative.

Color characteristics

A negative image, while not aesthetically pleasing to look at, nevertheless is a viewable image. A photographic negative film, then, like a photographic transparency film, is a complete imaging system. Each of the three basic imaging system functions of image capture, signal processing, and image formation takes place within the film itself (Figure 8.13).

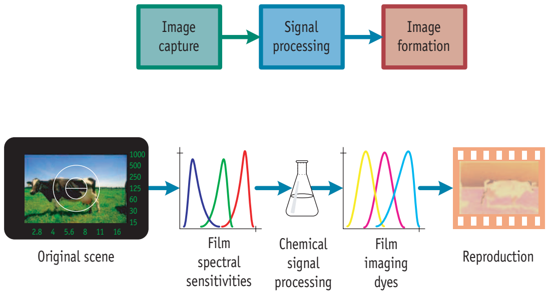


Figure 8.13 A photographic negative film is a complete, single-stage color-imaging system.

In addition, a negative film comprises the first *half* of a two-stage imaging system composed of the negative film and a print medium (Figure 8.14). In this and any other two-stage imaging system, the output image formed by the first stage becomes the input to the second stage.

It is often useful to analyze an imaging system having two or more stages in terms of a single *compound* system (Figure 8.15). If that is done for this example, it can be seen that the spectral characteristics of the *image-capture* stage of the compound system are determined solely by the red, green, and blue spectral sensitivities of the negative film. It also can be seen that the spectral characteristics of the *image-formation* stage are determined solely by the cyan, magenta, and yellow image-forming dyes of the print medium. Everything else within the compound system can be classified as *signal processing*. In other words, everything else serves to influence the *output* of the compound system (how much of each image dye forms on the print medium) as a function of system *input* (the amounts of red, green, and blue exposure recorded at the original scene by the negative film).

The total signal processing of the compound system includes chemical signal processing that occurs within the photographic negative film and additional chemical signal processing that occurs within the print medium. In addition, the *transfer of information* between the two stages—the way the print-medium stage "sees" the color information recorded on the negative—is a form of color signal processing. The signal processing consequences of the transfer process are illustrated in Figures 8.16 and 8.17.

Figure 8.16 shows an idealized example of the information transfer. In this example, the image-forming dyes of the negative film are hypothetical "block dyes." The spectral responsivities of the print medium are of very narrow bandwidth, they are completely separated from each other, and they are perfectly aligned with the spectral absorptions of the corresponding negative dyes. As a result, the cyan dye of the negative uniquely controls the red-light exposure to the print medium. Therefore, that cyan dye has only red printing density and has no green or blue printing density.

Similarly, the magenta dye of this idealized example uniquely controls green-light exposure to the print medium, so it has only green printing density and has no red or blue printing density; and the yellow dye uniquely controls blue-light exposure to the print medium, so it has only blue printing density and has no red or green printing density. In this

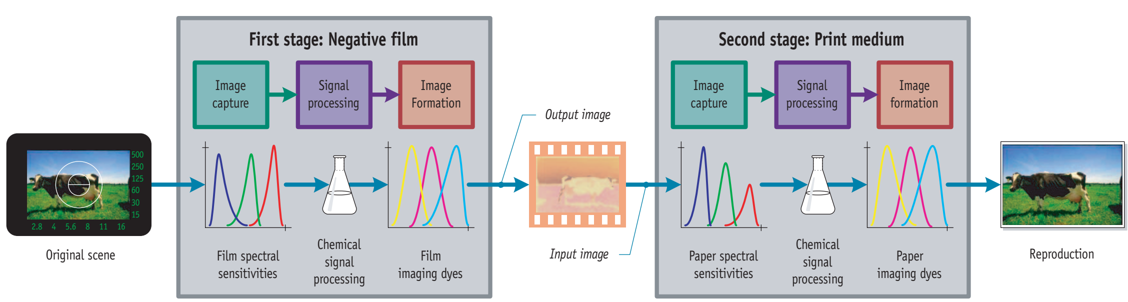


Figure 8.14 A photographic negative film is the first stage of a two-stage imaging system. In multistage imaging systems, an output image formed by one stage then becomes the input image to the next.

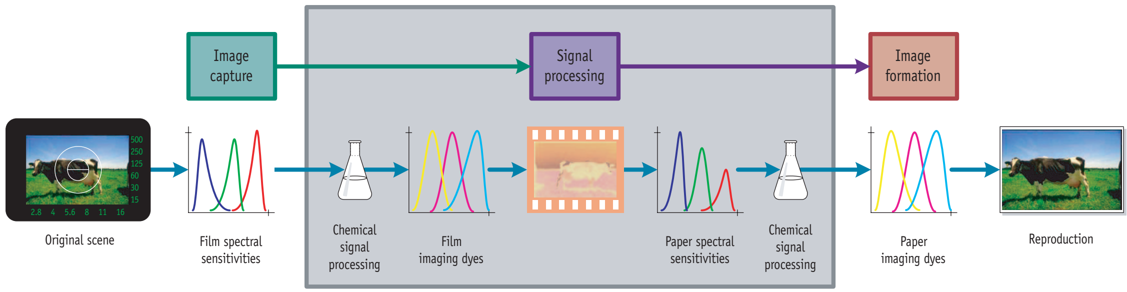


Figure 8.15 A two-stage imaging system can be analyzed in terms of a single compound system.

example, then, there is *no* crosstalk in the transfer of color information from the negative to the print medium.

Figure 8.17 shows a more realistic example. In this case, the cyan dye of the negative absorbs not only red light, but also some green and blue light (as "seen" according to the spectral responsivities of the print medium). The cyan dye therefore has *unwanted absorption* of green and blue light. In other words, it not only has red printing density, but also has green and blue printing density. The other dyes also have some unwanted absorptions. In particular, the magenta dye absorbs not only green light, but also some blue light. As a result of these unwanted absorptions, crosstalk is introduced in the transfer of color information from the negative to the print material.

This type of crosstalk must be counteracted by the use of appropriate color-correction mechanisms that create compensating crosstalk (interactions) among the color layers of the negative. Color correction can be provided by chemical interlayer effects, as is done in photographic transparency films. Because color-negative films are not viewed directly, another color-correction mechanism also can be used. It is this second mechanism—the incorporation of what are called *colored couplers*—that gives negative films their distinctive overall orange color cast. (Refer to Appendix C for more details on photographic media, chemical interlayer effects, and colored couplers.)

Color-encoding considerations

Several important issues related to color encoding have been uncovered in this analysis of color photographic negative films. First, the analysis has shown that the spectral sensitivities of color-negative and transparency films are not basically different, which means that color images on positive and negative photographic media are fundamentally equivalent *at the image-capture stage*. This is an extremely important finding; it will be shown later that it provides the key to successful *scene-based* color encoding.

The analysis also has shown that negatives having different sets of image-forming dyes can be metameric to a print medium, i.e., they can have the same printing densities. Printing-density metamerism has important implications for scanning and color encoding. Very few image scanners measure negatives in terms of printing densities; most measure values that are closer to Status M densities. As a result, *RGB* scanned values alone are not accurate predictors of printing-density metamerism. For example, areas on two different negative films might have identical printing densities, but their *RGB* scanned values might indicate that the two areas are different.

Negative films frequently are used under circumstances that lead to significant variations in overall exposure. For example, they often are used in cameras having no automatic or manual exposure control. The resulting exposure errors produce negative images that are lighter or darker than they would be if properly exposed. In addition, negative films typically are used under a wide variety of scene illumination conditions. This results in considerable color-balance and density variations among negative-film images, even among images on the same roll of film.

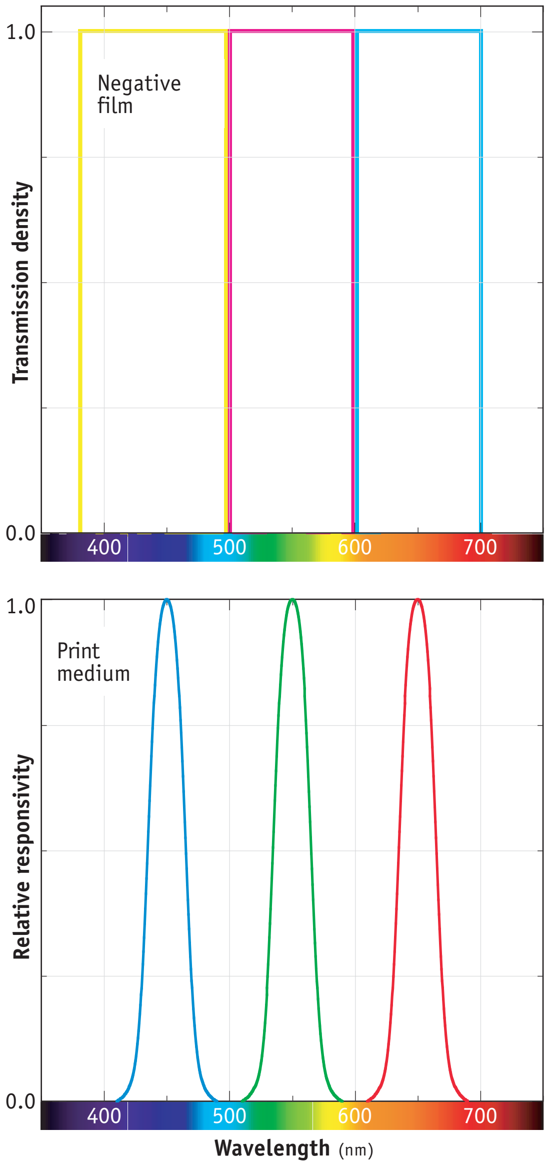


Figure 8.16 There is no crosstalk in the transfer of information from a negative to a print medium if the spectral absorptions of each image-forming dye in the negative uniquely align with the corresponding spectral responsivities of the print medium.

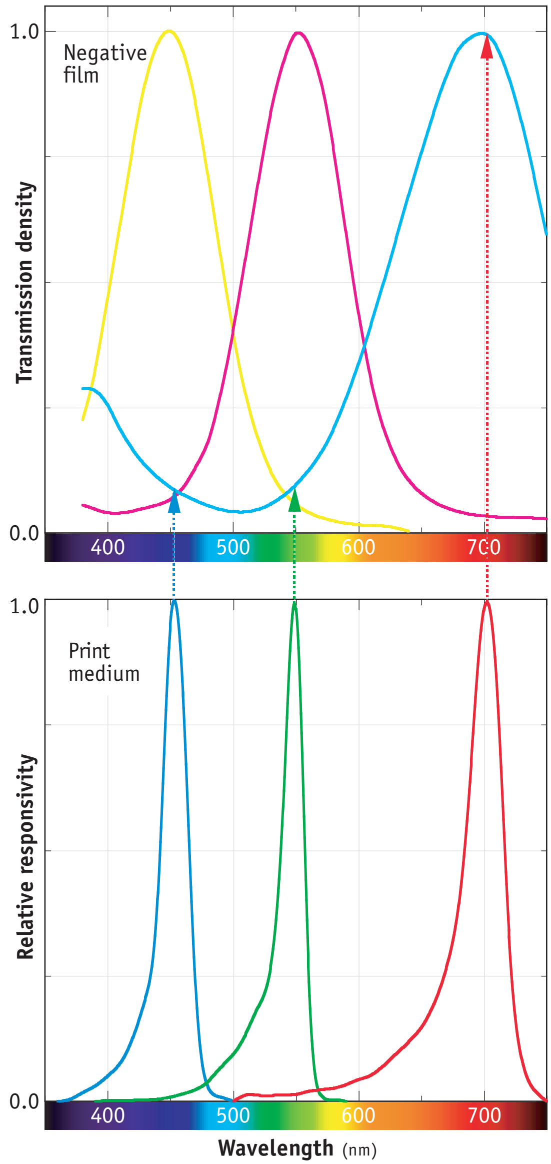


Figure 8.17 In most real cases, there is some crosstalk in the transfer of information from a negative to a print medium. For example, the cyan dye of the negative will modulate not only red light, but also some green and blue light as "seen" by the print medium.

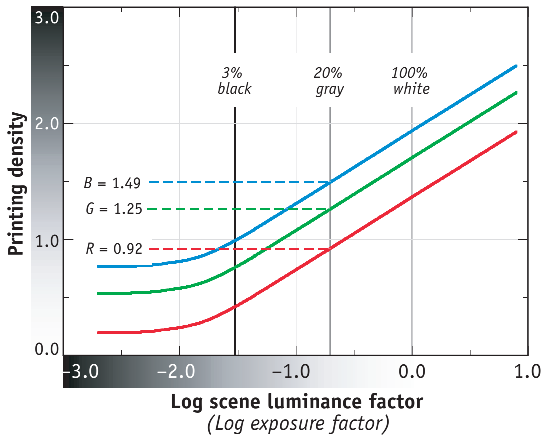


Figure 8.18a Placement of scene information on a normally exposed negative. The indicated *RGB* values correspond to a 20 % reflectance gray in the original scene.

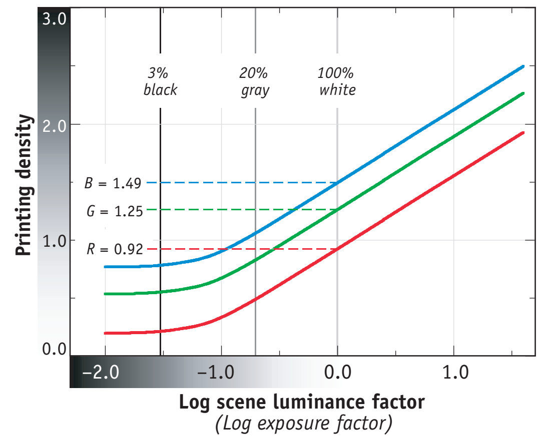


Figure 8.18b Placement of scene information on an underexposed negative. The same *RGB* values now correspond to a 100 % reflectance white in the original scene.

While such variations ordinarily are compensated for in the optical printing process, they introduce a basic *ambiguity* into the meaning of values scanned directly from negatives. For example, consider three negative images taken of the same scene. One image is normally exposed, one is somewhat underexposed, and one is considerably overexposed (Figures 8.18a, 8.18b, and 8.18c). As the figures show, the values $R = 0.92$, $G = 1.25$, and $B = 1.49$, measured by a film scanner, would represent a medium gray (20 % reflectance) in the original scene, if the negative being scanned were normally exposed. However, the *same* set of *RGB* values might represent a scene perfect white (100 % reflectance) if the negative were underexposed, or a scene dark black (3 % reflectance) if the negative were overexposed. There is a basic ambiguity, then, in any set of *RGB* values measured from a negative by a scanner.

Another challenge for digital imaging applications is that negative images, as mentioned in the chapter introduction, are *intermediary* images. Color values measured from the film itself do not correspond directly to original-scene colorimetric values, because the formation of a negative image involves a significant amount of chemical signal processing. On the other hand, color values measured from a negative film also do not correspond directly to colorimetric values of the image that ultimately will be produced from that negative, because those values will vary depending on the

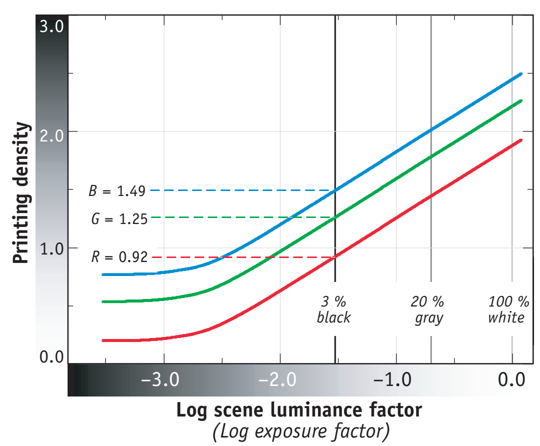


Figure 8.18c Placement of scene information on an overexposed negative. The same *RGB* values now correspond to a 3 % reflectance black in the original scene.

characteristics of the printer, print medium, and viewing illuminant.

All of this raises some serious problems for the color encoding of images originating on color-negative films. The color-encoding process will have to deal both with the ambiguity associated with color values measured directly from negatives and with the intermediary nature of color-negative images. This suggests that a successful color-encoding process is going to have to do more than just *measure* color values from negatives. Specifically, in order to represent color in a meaningful way, the encoding process will have to *interpret* color measurements.

In Part III, it will be shown that interpretation of measured values is required not just for input from negative images, but for images from *all* input media and from *all* input devices. It will be shown, in fact, that the concept of interpretation is the key to all successful digital color encoding.

Summary of key issues

- The low photographic gamma, extensive exposure dynamic range, and straight-line grayscale characteristic of color photographic negative films make them well suited for input to digital color-imaging systems.
- At the image-capture (latent image) stage, images on positive and negative photographic films essentially are the same.
- The neutral and color characteristics of all color-negative films are designed in terms of printing densities.
- Printing densities are the densities that the print material "sees" when it "looks" at the negative film, as illuminated by the printer light source.
- Negative films having different sets of image-forming dyes may have areas that are spectrally different but identical in terms of their printing densities. Such areas are metameric to the print material. If printed identically, they will produce the same color in the final print.
- Negative films are components of systems composed of two or more stages. The linkage between the stages—the way information recorded on the negative is transferred to another device or medium—is a form of color signal processing.
- Colored couplers are one mechanism used in color-negative films to compensate for certain unwanted absorptions of their image-forming dyes. They are responsible for the orange color cast of unexposed areas of the film.
- Negative films often are used under circumstances that lead to significant variations in exposure, which result in corresponding variations in density. Therefore density measurements of negative films, by themselves, are ambiguous.
- Negative films produce intermediary images. Color values measured from these images do not correspond directly to the colorimetry of the original scene, nor do they correspond directly to the colorimetry of the final print.
- Color measurements made directly from color-negative films must be interpreted appropriately in order to be meaningful.

PART III

Digital Color Encoding

Most discussions of digital color encoding focus on two areas: image file formats and color spaces. The underlying assumption of such discussions is that if the industry could define file format standards and then agree to use the same "device-independent" colorimetric color space(s), the problems of representing color in digital form and interchanging images among color-imaging systems would be solved.

The discussion presented in this part will be quite different, primarily because we do not concur with that basic assumption. Our experience instead has shown that the current problems of color encoding and interchanging digital images cannot be eliminated simply by file format standards and an industry-wide adoption of *any* standard colorimetric color space.

The basic reasons for this were introduced in Part II, where it was shown that different types of imaging media *do* have, and *must* have, fundamentally different colorimetric properties. Using the same colorimetric color space to encode images from these media, and doing nothing more, will in no way account for such differences. All it will do is quantify those differences in terms of the agreed-upon color space!

In this part, strategies will be devised to deal appropriately with the basic colorimetric differences that exist among various combinations of input imaging media and devices. First, some basic color-encoding concepts will be described. Several encoding methods that have been used successfully in commercial color-imaging systems then will be examined. These examinations will show why some current methods work well in certain situations and why they fail in others.

Much of the discussion in this part will focus on a concept that the authors refer to as *input compatibility*. We believe this concept to be the key to all successful color encoding. Also in this part, the concept of "device-independent color" will be critically examined.

Digital Color Management: Encoding Solutions, Second Edition Edward J. Giorgianni and Thomas E. Madden
 ISBN: 978-0-470-51244-9

9
Encoding Concepts

The basic function of digital color encoding in imaging systems is to provide a digital representation of colors for image processing, storage, and interchange among systems. Within a given system, encoding provides a digital link between the system's inputs and outputs.

In a simple system, having just one type of input and one type of output, color encoding can be performed prior to any signal processing (Figure 9.1). The encoding therefore is a direct representation of the color values measured by the system's input device.

In more complex systems, supporting multiple types of inputs and outputs, such an arrangement is impractical because each and every *combination* of input and output would require a separate signal processing transform. For example, a single output device would require two different transforms in order to process color values measured by one input device that scans photographic negative films and another that scans reflection prints. The number of required system transforms in this arrangement equals the *product* of the number of inputs and outputs. That can get to be a sizable number of transforms. A system having four inputs and eight outputs, for example, would require 32 different transforms (Figure 9.2).

A much more efficient system results if the color signal processing is split into two parts: input signal processing and output signal processing. In this arrangement, shown in Figure 9.3, each input and each output has only a single associated transform. The number of system transforms then equals the *sum*, rather than the product, of the number of inputs and outputs. Only 12 signal processing transforms, instead of 32, would be required for a system supporting four inputs and eight outputs.

This signal processing arrangement has other advantages that can best be described by looking at one very successful system in which it is used—the audio compact disc (CD) system (Figure 9.4a). This system is an excellent example of good multiple input/output design. One of the most important features of the overall system is that it supports input from a wide variety of sources. For example, a single CD may store digital information from sounds previously recorded on such different media as wax cylinders, lacquer discs, wire, vinyl discs, or tape. Some input sources, especially audio tapes, may have been recorded using either analog or digital technology, and any of a number of different noise suppression and frequency pre-emphasis techniques may have been included in the original recording process.

Digital Color Management: Encoding Solutions, Second Edition Edward J. Giorgianni and Thomas E. Madden
 ISBN: 978-0-470-51244-9

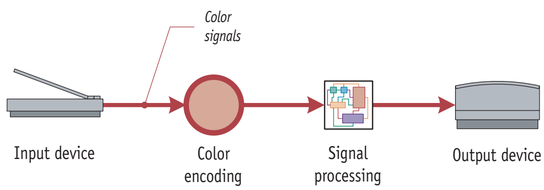


Figure 9.1 A simple color-imaging system.

The system works well, despite the disparities of the input media, because appropriate signal processing transforms are used on all input signals. As a result, data stored on an audio CD represent processed signals that have been digitally encoded and stored in an efficient and standardized way. Moreover, all stored data essentially are *independent* of the original audio source.

Without that independence, the original audio source would have to be identified, and the signal processing of the CD player would have to be altered accordingly (a situation essentially corresponding to Figure 9.2). This would greatly increase the cost and complexity of the CD player. It also would preclude the future addition of different types of input sources to the overall system, because existing players would not have the corresponding signal processing required to properly play back discs made from those sources.

Note, however, that the use of standardized encoding does not mean that the system *output* necessarily is fixed or restricted. Additional output signal processing—in the form of volume and tone controls, graphic equalizers, digital signal processors, etc.—can be used by the listener to optimize the output sound ultimately produced by the system (Figure 9.4b). Factors such as listener preferences

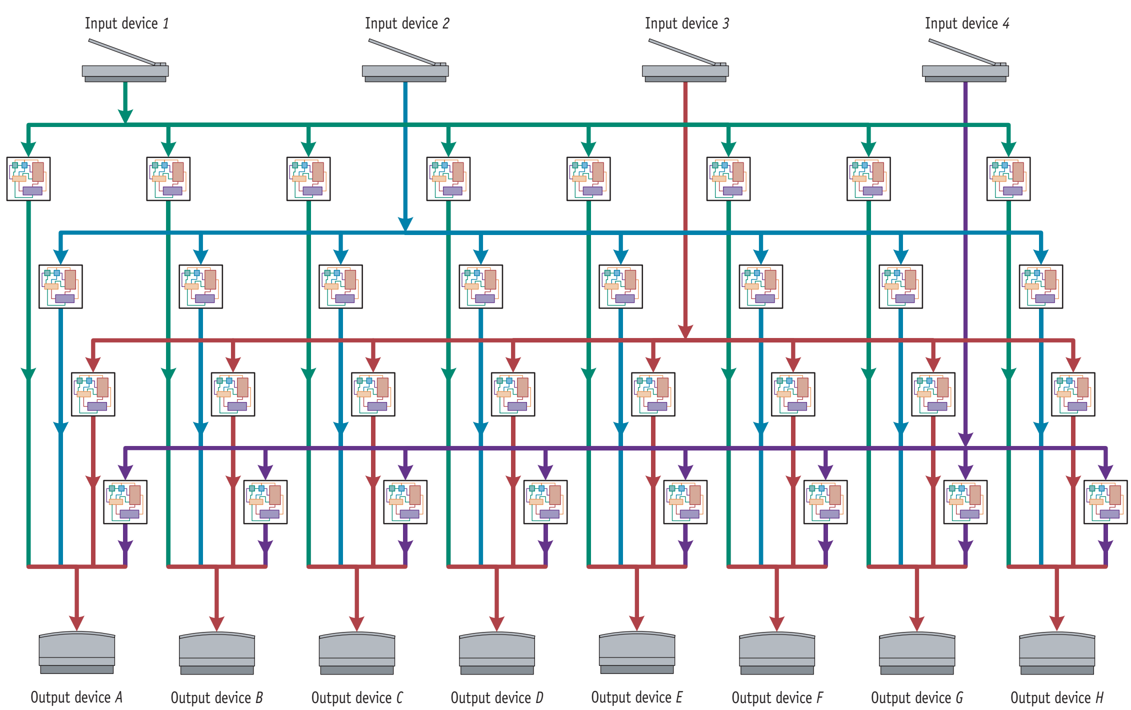


Figure 9.2 A system with four inputs and eight outputs. In this arrangement, 32 signal processing transforms are required.

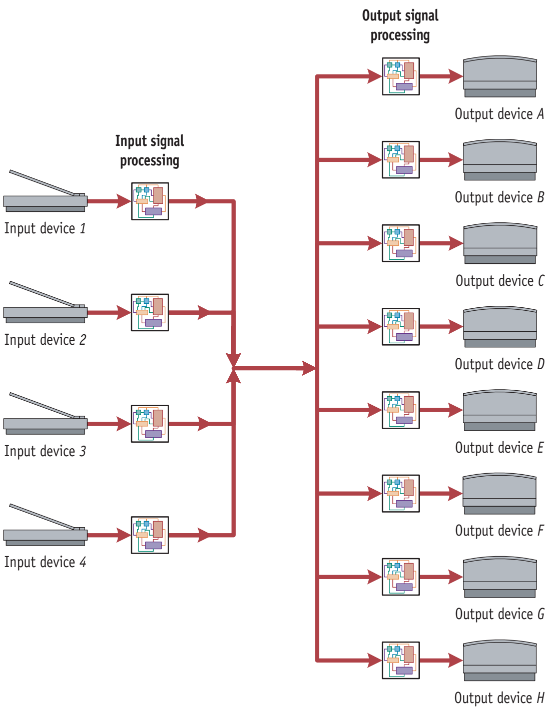


Figure 9.3 In this arrangement, only 12 signal processing transforms are required.

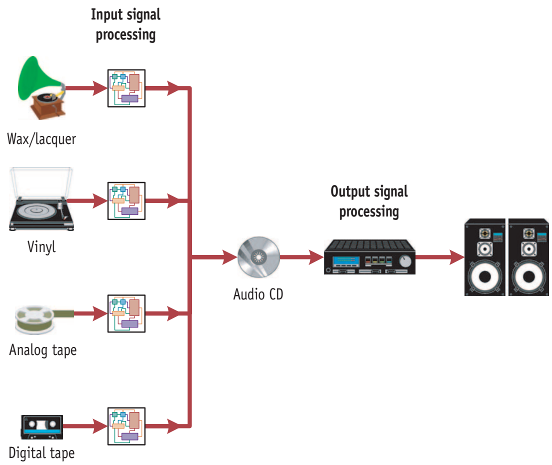


Figure 9.4a The audio CD system is an excellent example of a well-designed multiple input/output system.

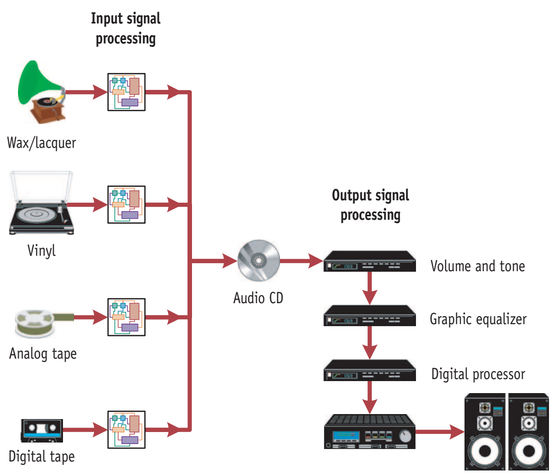


Figure 9.4b Some output signal processing options in an audio CD system.

and listening environment characteristics will affect this optimization process.

From this examination, it can be seen that the audio CD system incorporates a number of important features: the system provides for input from a wide variety of different sources; it allows for future incorporation of additional input sources; it has a means for encoding, storing, and exchanging data in an efficient and standardized way; its basic output signal processing is independent of the original input source; additional output signal processing can be used to optimize the final output of the system; and it allows for future incorporation of additional output sources.

All of these features would be very desirable in a multiple input/output color-imaging system. Providing them requires that a suitable signal processing arrangement be used and that careful attention be paid to the encoding of color, as discussed next.

Multiple input/output color-imaging systems

Figure 9.5 shows a signal processing arrangement for a multiple input/output color-imaging system. As in the audio CD system, this arrangement uses a

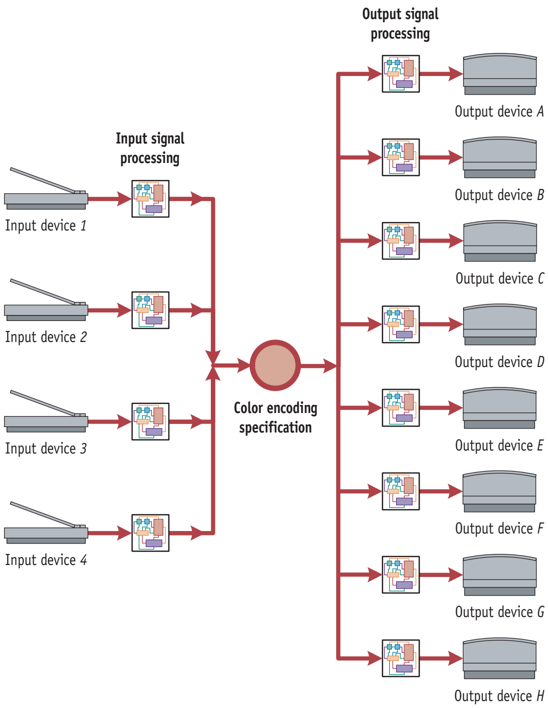


Figure 9.5 Use of a color encoding specification (CES) in a multiple input/output color-imaging system.

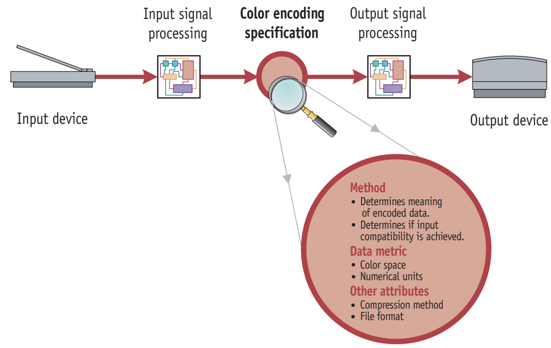


Figure 9.6 A color encoding specification must define two principal attributes: a color-encoding method and a color-encoding data metric. Other attributes also may be defined.

standardized input/output interface at the center of the system. The interface must provide a standardized representation of color, defined in terms of what we will refer to as a *color encoding specification (CES)*.

A color encoding specification must define two principal attributes of the color representation: a *color-encoding method* and a *color-encoding data metric* (Figure 9.6). The distinction between these attributes is extremely important.

The color-encoding method determines the actual *meaning* of the encoded data, while the color-encoding data metric defines the *color space* and the *numerical units* in which encoded data are expressed. For example, the encoding method might be such that the color data represent standard CIE colorimetric measurements of reproduced images. (Whether that method will work or not will be discussed later.) Those measurements then can be expressed in terms of various *data metrics* without changing their basic meaning. For example, any of several different *color spaces*, such as CIELAB or CIELUV, could be used. Color-space values in turn can be digitally encoded in various ways, again without changing the basic meaning of the data. For example, CIELAB values might be scaled in a particular way and represented by three 8-bit numbers (Figure 9.7).

The selection of a data metric is one of many important engineering decisions that must be made for any color-imaging system. The choice will affect the

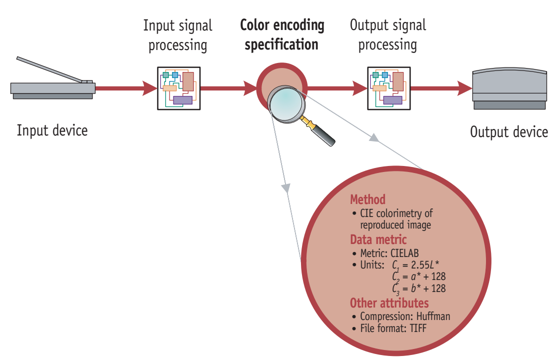


Figure 9.7 An example of a color encoding specification.

efficiency of the color encoding, and it will directly affect other factors such as image compression, as will be shown in Chapter 13. But the color-encoding method is much more fundamental. It defines the entire basis for the standardized representation of color within the system. The critical point is this: *If the color-encoding method that is selected is inappropriate, the system will fail, regardless of which data metric is chosen.*

The reason for this is that the encoding method determines *what will be represented* by the encoding, whereas the data metric determines only *how the representation will be numerically expressed*. As an analogy, the selection of a method is equivalent to determining whether something will be measured in terms of its volume, weight, or mass. Once that is determined, various data metrics then can be used to express the measurement without changing its basic meaning. For example, the method might determine that volume will be measured. That measurement then can be expressed in cubic inches, cubic centimeters, or various other units without changing its meaning.

In addition to a color-encoding method and a color-encoding data metric, a color encoding specification also may include other attributes, such as a data compression method and a data file format. Although they are not strictly part of the color representation, those attributes also must be completely defined so that images encoded according to the specification can be interchanged among various imaging systems and applications. Some of those attributes, as they relate to color, will be discussed in later chapters. For now, our focus will be on color-encoding methods.

The concept of input compatibility

The most effective color encoding specification for a given system will depend on the specific requirements for that particular system. Nevertheless, there is one rule that *always* must be obeyed: to be successful, a color encoding specification must be based on a color-encoding method that creates a condition we refer to as *input compatibility.*

A color-imaging system in which the inputs have been made compatible will have a number of important features, similar to those of the audio CD system:

- Each output device can produce images from any image encoded in terms of the color encoding specification, regardless of the type of input medium or device from which the encoded image data were produced.
- Image data encoded from one input medium can be intermixed with image data encoded from another. For example, image data encoded from two different types of input media can be merged (cut and pasted) to produce a seamless composite image.
- The number of signal processing transforms required by the system is minimized.
- New input sources can be added to the system without requiring the addition of new corresponding output transforms. Each new input source requires only a transform relating that input to the color encoding specification.
- Similarly, new output devices and media can be added to the system without requiring the addition of new corresponding input transforms. Each new output requires only a transform relating the color encoding specification to that output.

Creating input compatibility requires that—at a minimum—all fundamental color differences, such as those that exist between negative and positive input media, be eliminated as part of the encoding process. In addition, image data from all input sources must be encoded on some *common basis*.

When this is done appropriately, interpretation of encoded color values does *not* require knowledge of the input image medium that was the source of that color, nor does it require knowledge of the device that was used to measure the color. Encoded input-compatible values *alone*, expressed in terms of the color encoding specification, completely and unambiguously define color.

For example, a triad of input-compatible encoded values, such as 83, 105, and 52, expressed in a particular data metric, would be sufficient to unambiguously define a color. If some *additional* qualification is necessary in order for the color to be completely defined—such as "the values were measured from (a particular medium)" or "the values were measured by (a particular scanner)" or "the values are meant for (a particular output device)"—then the meaning of the values alone must be ambiguous. So a *negative* test for input compatibility can be defined: if there is any ambiguity or inconsistency in the encoded values, then a condition of input compatibility has *not* been achieved. This test will be used often in the upcoming discussions of color-encoding methods.

Creating input compatibility is not always easy. In particular, it can be difficult to establish a common basis for color encoding in systems where the inputs are of fundamentally different types, such as systems that support input from both positive and negative photographic films and systems that support input from both hardcopy and electronic sources. Nevertheless, achieving input compatibility is the key to successful color encoding, and the focus of the next chapters will be on achieving that compatibility.

Input compatibility and tags

It is often suggested that the features we have associated with input compatibility can be achieved instead by the use of image file *tags*. Tags are image file headers containing additional information about the image data in the file. Such information is useful for specifying basic properties, such as image size and spatial resolution. Tags also can be used to provide information required to interpret the encoded color data. This type of identification is the basis for file format standards such as *TIFF* (Tagged Image File Format).

Although the use of tagged files certainly is a vast improvement over using image files in which there is no way of knowing what the numbers really mean, tagging alone is not a *substitute* for achieving input compatibility. The previously described system features cannot be achieved unless the fundamental differences among the various inputs are removed somewhere along the signal processing chain. For example, mixed image types, even if tagged, cannot be *directly* edited into seamless composites. Editing together portions of an image from a negative film with portions from a positive film certainly would not be seamless unless the image data first are translated into some common (input-compatible) form.

The use of tags does, however, provide flexibility as to *when* and *where* input compatibility is created. For example, in Figure 9.8a, input compatibility is achieved in the color encoding specification itself. In Figure 9.8b, image files first are produced, in any of a variety of different color encoding specifications, and tagged appropriately. Input compatibility is not created until later, when the tagged image files are read and processed on a workstation.

Both approaches are valid, and the decision as to which should be used for a given system will depend on the particular capabilities and requirements of that system. In most circumstances, however, we favor the approach in which input compatibility is created in the color encoding specification itself. Systems based on that approach still can use tags—all of the systems that we have designed have, in fact, used tags. But such systems are not dependent on

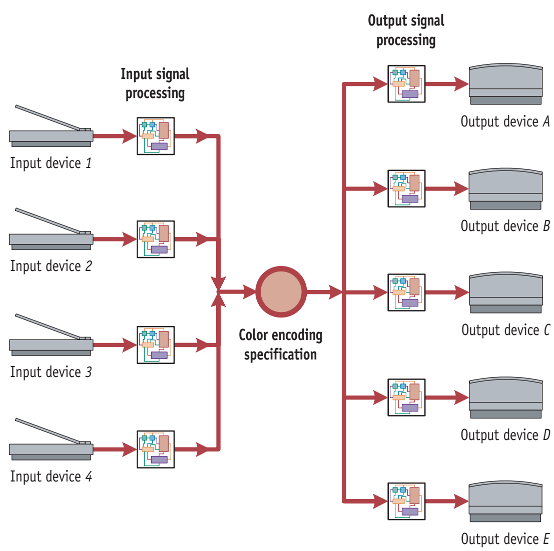


Figure 9.8a To achieve the system features described in the text, input compatibility must be created somewhere along the signal processing chain. One option is to create that compatibility directly in the color encoding specification.

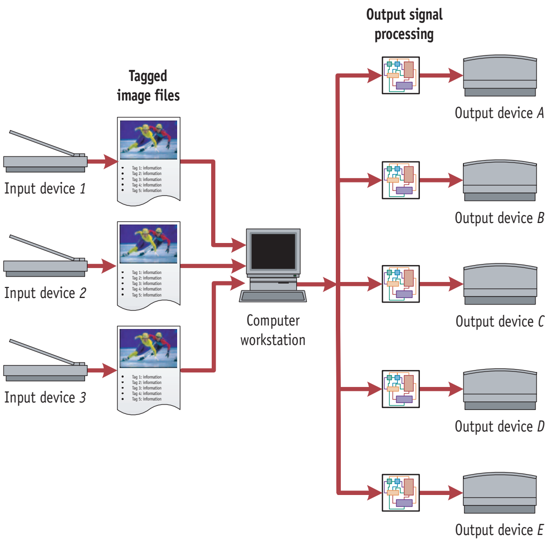


Figure 9.8b If image files are tagged appropriately, input compatibility can be created elsewhere, such as in a color-management application running on a workstation.

tags for their basic color behavior. Tags instead are used to provide supplemental information that can enhance overall system performance. For example, a sophisticated user or application might perform some special image processing if auxiliary information about the origin and destination of an image is known.

One concern regarding an approach that is entirely dependent on the use of tags to define the meaning of encoded data is that the problems caused by different input types are not solved up front; they are passed to the user of the workstation. That may be perfectly acceptable if that user is knowledgeable and/or has access to applications that can deal appropriately with the tagged information, but that may not always be the case.

That concern is eliminated if any input compatibility problems are solved in the color encoding specification, before image files are created. When that is done, simpler applications and relatively unsophisticated users can more easily make use of the image files that are subsequently produced. An additional concern is that, in practice, image file tags have a way of getting lost or corrupted as images are passed from system to system, edited, or merged with other images. Tagged information also may never have been filled in, or information may have been entered incorrectly.

In our opinion, then, whenever it is possible to do so, a system should be designed to work well with all image files, even if any supplemental information that might be supplied by tags is missing. In some systems, there really is no choice, because image files are meant to be played "as is" on an output device. Under these circumstances, when there is no opportunity for image data to be processed on an intermediary workstation or other signal processing device, input compatibility must be created in the color encoding itself.

Input compatibility and device-independent color

Another frequent suggestion is that input compatibility can be achieved simply by the use of a so-called "device-independent" color metric to encode data from all input sources. The most common proposals put forward for "device-independent" color encoding advocate using metrics based on a CIE color space such as *XYZ*, CIELAB, or CIELUV. However, the use of such metrics *alone*, i.e., without regard to the encoding *method*, will *not* ensure input compatibility.

One way to explain the distinction between the use of a device-independent encoding metric and the achievement of true input compatibility is by analogy. Suppose that instead of the scanning and reproduction of images, the problem was the scanning and reprinting of pages of printed text. If such a system is limited to a single input and output language—if, for example, it is used for scanning and reprinting English words only—the encoding can be simple (Figure 9.9).

Because the text-scanning system shown in Figure 9.9 is restricted to a single input/output language, the encoded information only would have to represent the patterns of black and white on the scanned input page. Under these circumstances, even a conventional photocopier would be sufficient.

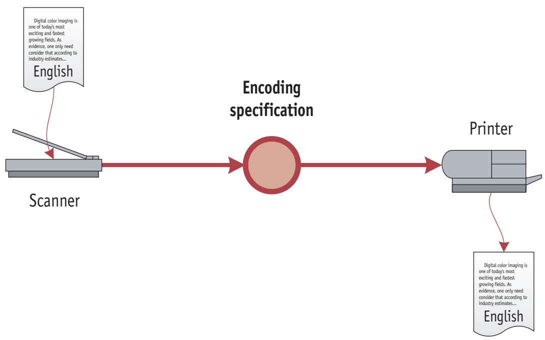


Figure 9.9 A simple text-scanning system, which is restricted to a single input/output language.

Now suppose that there was an additional requirement for the system also to scan text that is written in two other languages, say French and Spanish (Figure 9.10). Again, the encoding can be simple, and a conventional photocopier will be sufficient, *if* the system is restricted to reprinting only English from input English texts, reprinting only French from input French texts, and reprinting only Spanish from input Spanish texts.

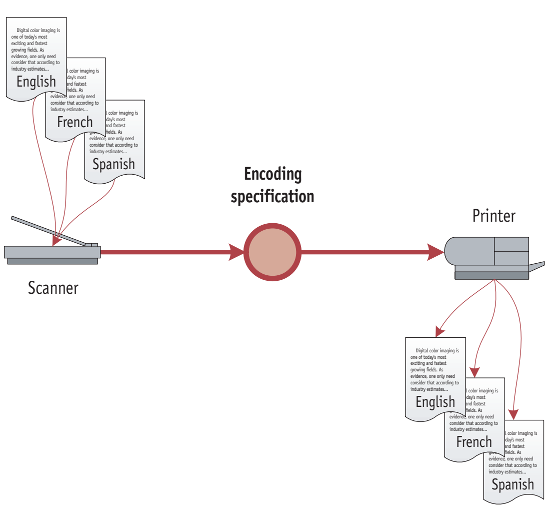


Figure 9.10 A multiple language text-scanning system. This system is restricted in that the output must be in the same language as the input.

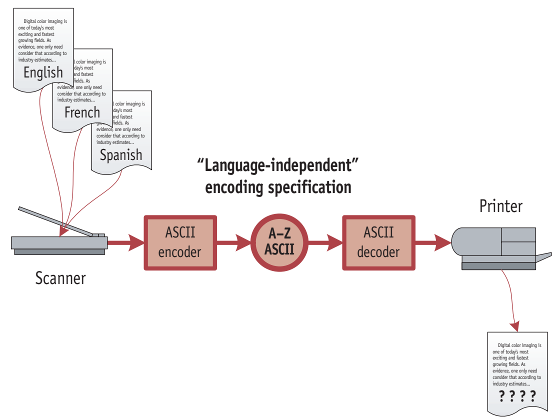


Figure 9.11 A text-scanning system that is based on a common "language-independent" metric. The system will not work.

But what if the system was required to scan words in *any* of the input languages and print out these words in *any* of the output languages? And what if, in addition, the system was to be used to edit input text? In particular, what if it was to be used to cut and paste words scanned from different languages and then to produce composite pages of text in any of the output languages? Would an approach analogous to the use of a "device-independent" color metric work?

Let us try using a "language-independent" text metric to see what happens. The Roman letters A–Z can serve as such a metric because they can be used to represent text in any of the input languages. The scanned letters can be encoded in ASCII, which provides a convenient, efficient, standardized language-independent encoding metric (Figure 9.11).

But that language-independent encoding would contribute *nothing* toward solving the basic problem. The reason is that the encoding alone would be *ambiguous* (the key word in our test of input compatibility). The same pattern of encoded letters could have entirely different meanings, depending on what input language is assumed. For example, the meaning of the pattern of letters p-l-u-m-e is ambiguous (Figure 9.12). That pattern means "pen" in French, but it also means "large fluffy feather" in English!

So despite the use of a language-independent text-encoding metric in this example, the *meaning* of a

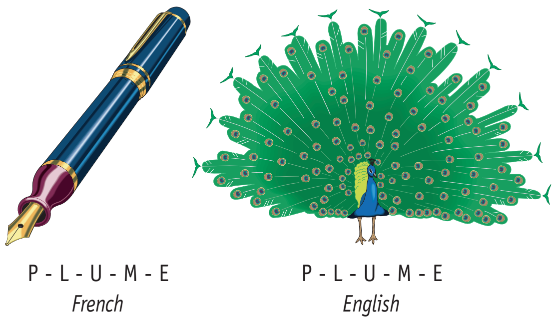


Figure 9.12 The meaning of the pattern of letters "P-L-U-M-E" is ambiguous!

given pattern of letters remains *language dependent*. The language of origination would have to be identified, perhaps by a tag, along with the encoded letters in order to produce meaningful text output from the system.

The real problem, then, is that the letters alone are not sufficient to convey their meaning. It is important to understand that this is *not* a metric problem, i.e., it is not a problem with the A–Z characters or with ASCII. There is an *inherent* ambiguity, and therefore a basic incompatibility, in the inputs because they are in *different languages*. Since that ambiguity is not *caused* by a metric, it cannot be *solved* by the selection of a particular metric, whether that metric is language independent or not.

In order to achieve true input compatibility in this system, it is necessary to do more than convert all of the input words to a common *metric*. It is necessary to use an encoding *method* that *interprets the meaning* of all inputs and expresses that meaning in a *common language*.

Let us now try that approach to see what happens. It will not really matter what language is used as the common language (at least not yet), so let us use Latin. The system then would be as shown in Figure 9.13.

The role of each input transform in the arrangement now is apparent: each must *translate* from the particular input language to Latin. The role of each output transform also is apparent: each must *translate* from Latin to the particular output language. The output transforms are independent of the input language; it does not matter what language(s) the original input words were in, because once transformed to the encoding specification, *all* words are in Latin. The system can be used to cut and paste encoded words, again because they all are expressed in Latin. Editing can be done without knowing in

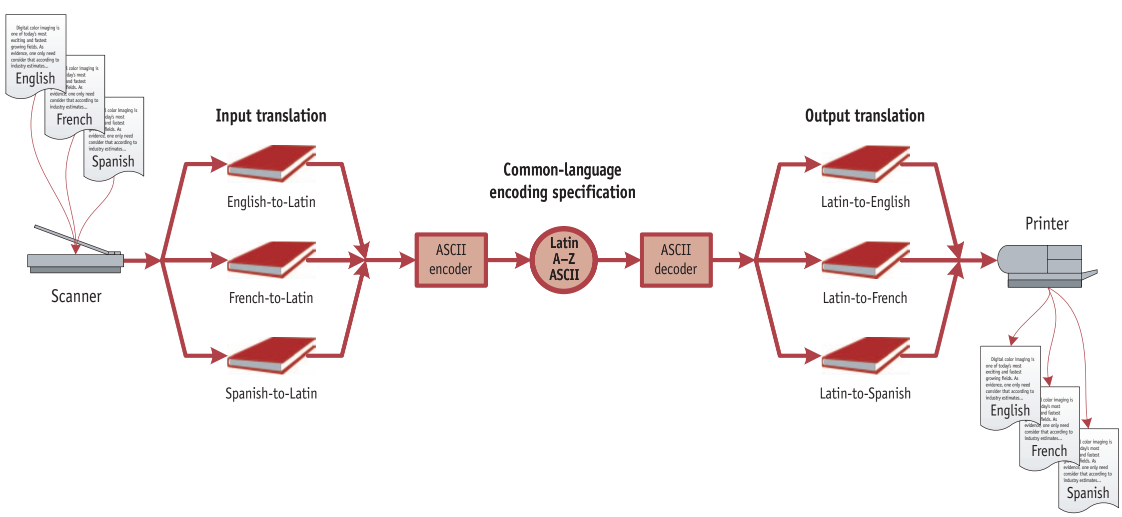


Figure 9.13 A text-scanning system based on a common encoding language. This system will work.

which language the final document will be printed. The system can, in fact, produce output in any or all of the output languages from the same encoded text. New input and output languages easily can be added; each would require just one translation transform. All of this was accomplished by the creation of *compatibility* among the system inputs; it could *not* have been done by the use of a "language-independent" metric alone.

Creating input compatibility in color-imaging systems

The language analogy essentially defines what is needed to establish input compatibility in a color-imaging system: *The meaning of the color information from each input source must be* interpreted *before encoding.*

That is not always simple to do. Consider the multiple input system shown in Figure 9.14. Color information provided by the digital still camera most likely was captured directly from live original scenes. Color information from the reflection and slide scanners comes from positive reproductions on two very different types of imaging media. And, to keep things interesting, the negative scanner provides information scanned from images that have recorded color "backwards."

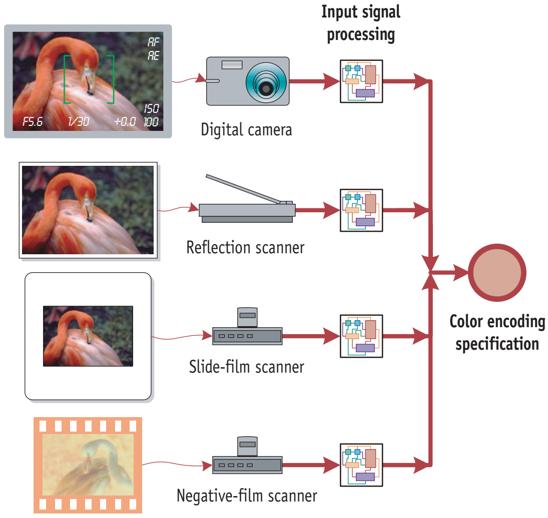


Figure 9.14 A color-imaging system with four fundamentally different types of inputs.

The system is dealing with inputs having *fundamentally* different colorimetric properties. These inputs are, in effect, speaking very different "languages." An appropriate color-encoding *method* must be used that will allow the *meaning* of the color information from each input to be interpreted and translated into a common color "language."

This translation, and the encoding method itself, must be based on some color property—a particular aspect of color—that all of the inputs have in common. It is that aspect of color that then must be measured and digitally encoded in order to represent color completely and unambiguously in the encoding specification. Color measurement and color-encoding methods that might be considered include the following:

Densitometric encoding:

- Measurements made according to spectral responsivities that are not equivalent to a set of visual color-matching functions.
- Encoding based on the densitometric measurements.

Standard colorimetric encoding:

- Measurements made according to the spectral responses of the CIE Standard Colorimetric Observer.
- Encoding based on standard colorimetric computations.

Advanced colorimetric encoding:

- Measurements made according to the spectral responses of the CIE Standard Colorimetric Observer.
- Encoding includes colorimetric adjustments for certain physical and perceptual factors.

In the next chapters, each of these measurement and encoding methods will be examined. These examinations will show which methods will work, and which ones will not, in various types of practical imaging systems.

Summary of key issues

- The basic functions of digital color encoding in a color-imaging system are to link system inputs and outputs and to provide a digital representation of color for editing, storage, and interchange.
- Support of multiple input and output types can best be accomplished by first dividing the digital signal processing into two parts—input processing and output processing—and by then using an appropriate color encoding specification to link the inputs and outputs.
- A complete definition for a color encoding specification includes both a color-encoding method and a color-encoding data metric. Other attributes, such as image file format, also may be defined.
- A color-encoding method defines the actual meaning of the encoded data, while a color-encoding data metric defines the color space and numerical units in which encoded data are expressed.
- Input signal processing associated with the color encoding specification must be based on an encoding method that creates input compatibility among all input sources.
- Input compatibility means that data from all input sources are encoded on a common basis and are expressed, in terms of a color encoding specification, such that encoded values completely and unambiguously define color.
- To achieve input compatibility, it is necessary to define the color encoding specification in terms of an aspect of color that all the system inputs have in common.
- When the colorimetric properties of the input sources are fundamentally different, appropriate input signal processing must be used to create input compatibility. In such cases, input compatibility cannot be achieved simply by the use of a common sdata metric.

10

Densitometric Color Encoding

Densitometric color encoding is based on input-image color measurements made according to defined sets of spectral responsivities that are not equivalent to a set of visual color-matching functions. The responsivities can be those of a particular type of densitometric instrument, such as an ISO Status A or Status M densitometer (Figure 10.1). The responsivities also can be those of an actual scanner or of some hypothetical reference scanner. Encoded colors can be expressed in terms of red, green, and blue densities, transmittances, reflectances, CMY or CMYK colorant amounts, or other values associated with the densitometric measurements.

The principal advantage of this type of encoding is that, because it corresponds quite directly to physical measurements of input images, it often simplifies color signal processing. Consider, for example, the system shown in Figure 10.2. Let us assume that the objective of the system is to produce output images that visually match scanned input images, and also that the output thermal print medium has CMY image-forming dyes that are *identical* to those of the input medium.

The responsivities of the system's densitometric scanner can be designed such that each R, G, or B channel primarily "sees" just one image-forming dye of the input medium (Figure 10.3). Each scanner color signal therefore will represent the amount of the corresponding C, M, or Y dye for each pixel in the scanned image. If the output is also designed such that each R, G, or B color signal primarily controls the amount of just one of the C, M, or Y image-forming dyes produced on the output medium, little or no color signal processing will be required in the system.

The system shown in Figure 10.3 simply will measure input-image dye amounts and produce those same dye amounts on the output. The output image, then, will be a one-to-one *physical* copy—a *duplicate*—of the input image. If the input and output images are viewed under identical conditions, they of course will *visually* match.

This example certainly is a special case. Nevertheless, it illustrates an important point: it is commonly assumed that output devices *must* be sent color-image data expressed in terms of human-observer-based colorimetry, such as CIE colorimetry, in order to produce a visual match to the input. But that is not true. Ultimately, output devices must generate color signals that will *result* in a visual match (or some other desired colorimetric result). But that does not necessarily mean that an output device must be *supplied* with CIE colorimetric values in order to do so. Nor does it mean that it is even useful to *measure* the CIE colorimetry of the input in all cases.

For example, in a system like that shown in Figure 10.3, a *colorimetric scanner* having responsivities equivalent to those of the CIE Standard Observer could be used instead of a densitometric scanner (Figure 10.4). The colorimetric scanner would provide direct measurements of CIE *XYZ* tristimulus

Digital Color Management: Encoding Solutions, Second Edition Edward J. Giorgianni and Thomas E. Madden

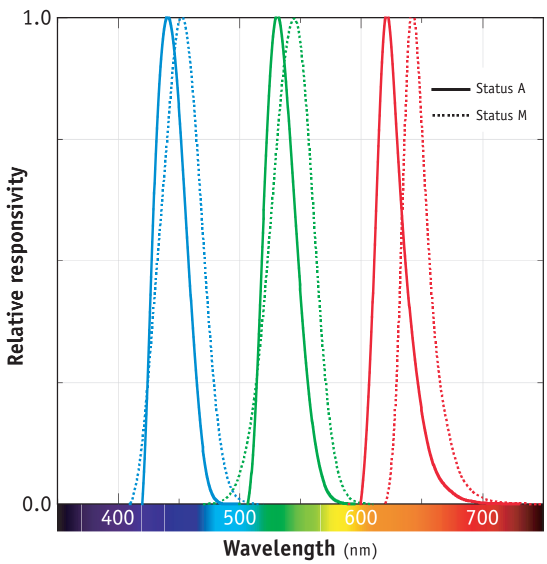


Figure 10.1 Two sets of red, green, and blue responsivities that could be used for densitometric color encoding: ISO Status A (solid lines) and Status M (dotted lines).

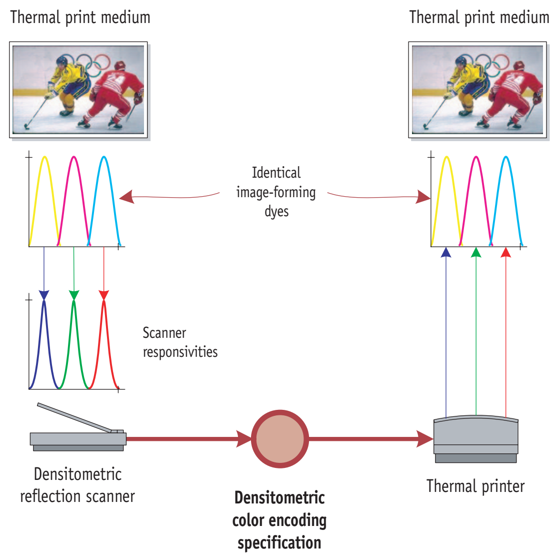


Figure 10.3 In this system, densitometric scanning and color encoding will minimize, and may even eliminate, the need for color signal processing.

values of the input image. But what would be done with those values? To produce an output image having those colorimetric values, it would be necessary to transform the *XYZ* values back to *RGB* values in order to produce the proper output-device signals. Unless there were some other purpose for the *XYZ* values (such as for image editing in a CIE color space), it would be far simpler to scan using a

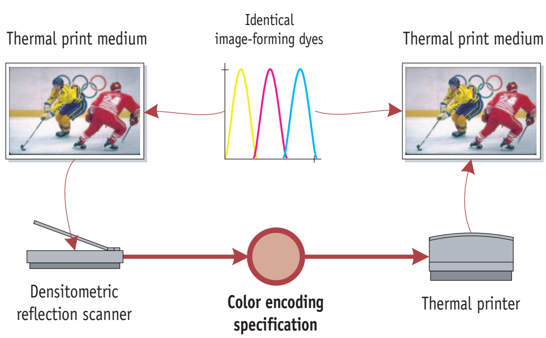


Figure 10.2 An imaging system using the same medium for input and output.

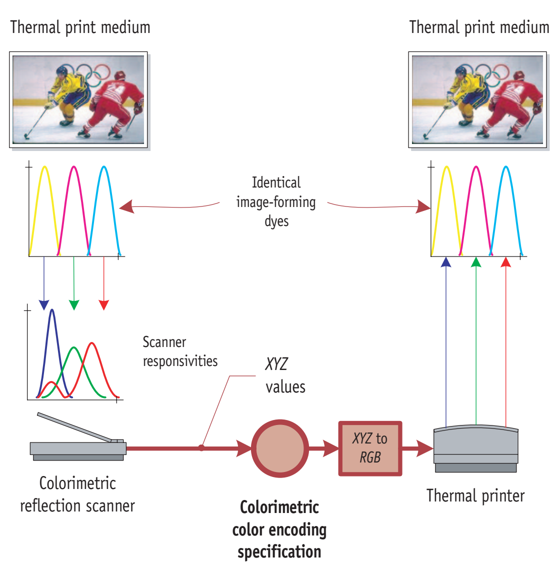


Figure 10.4 In this system, colorimetric scanning adds no real value, and its use requires additional color signal processing.

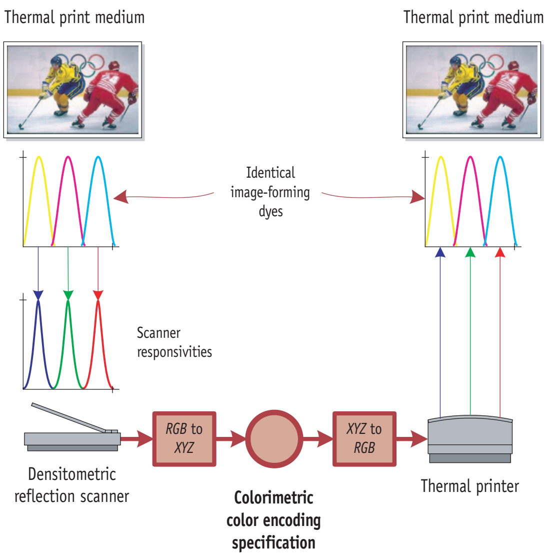


Figure 10.5 In this system, colorimetric encoding adds no real value, and its use requires additional color signal processing.

densitometric scanner and avoid the need for an *XYZ*-to-*RGB* output transform.

Also, in a system based on densitometric scanning of a single medium, an input transform could be used to convert scanner *RGB* values to CIE *XYZ* tristimulus values (Figure 10.5). But again, what would be done with those values? In order to produce an output print having that colorimetry, it again would be necessary to use an output transform. For the system shown, employing an input transformation from scanner *RGB* values to *XYZ* values followed by an output transform from *XYZ* back to *RGB* values serves no real purpose. Moreover, the use of unnecessary transformations may result in degradations in signal processing speed and accuracy.

Now, relatively few systems use the same imaging medium for both input and output. Still, the scanned *RGB* input and required *RGB* output values of many practical systems often are remarkably similar. They are, at least, much more similar to each other than either is to any set of CIE colorimetric values. Numerous systems have failed because of the unnecessary use of color transformations of scanned *RGB* values to colorimetric and other types of color spaces, followed by complex transformations from those spaces to *RGB* output values. In virtually every case, the systems would have worked much better if simple *RGB*-input to *RGB*-output transformations had been used instead.

Densitometric encoding and input compatibility

Densitometric color encoding generally can be used when an imaging system is restricted to having just one input medium and a single scanner. In such cases, there is no real issue of input compatibility because all scanned data inherently are input compatible. There are no fundamental disparities among the scanned images because all image data are produced from the same scanner measuring images on the same medium. The meaning of the scanned values themselves therefore is unambiguous.

What happens, however, if there is more than one scanner (but still just one input medium)? If all the scanners have identical densitometric characteristics, then the situation remains the same and densitometric color encoding will be sufficient (Figure 10.6). There will be no ambiguity, because an image measured on any of the scanners would yield the same scanned values.

But what if the scanners have *different* densitometric characteristics, or if the densitometric characteristics of a single scanner change over time? Because

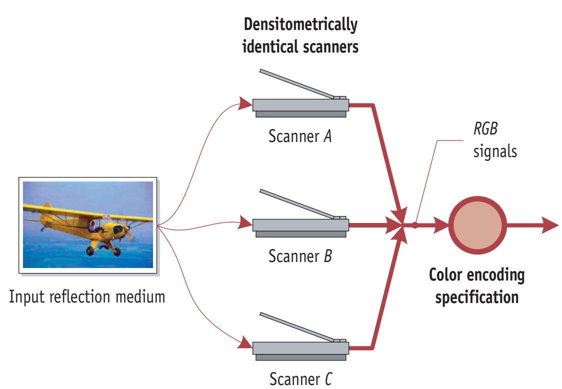


Figure 10.6 The use of multiple input scanners is not a complication if all scanners are identical.

there still is only one input medium, densitometric measurements again can serve as the basis for color encoding. However, input compatibility will not result automatically from a direct encoding of these measurements. The meaning of measured values alone will be ambiguous; it will depend on which particular scanner is used (or on when the values are measured by a scanner having densitometric characteristics that change over time).

One way of dealing with a multiple scanner situation would be to equip each output device with an arsenal of output transforms—one for each input scanner. A simpler solution, however, would be to create input compatibility among the scanners themselves. The methodology for doing so also addresses the problem of scanners that change in their densitometric characteristics over time.

In order to create input compatibility among multiple (or variable) scanners, the signal processing of each scanner in the system must include an appropriate correction transform, based on a *calibration* of the particular device (Figures 10.7a and 10.7b). The correction transforms primarily would compensate for differences in the spectral responsivities among the scanners. The transforms also would correct for other measurement differences due to the optical and electronic properties of the individual scanners.

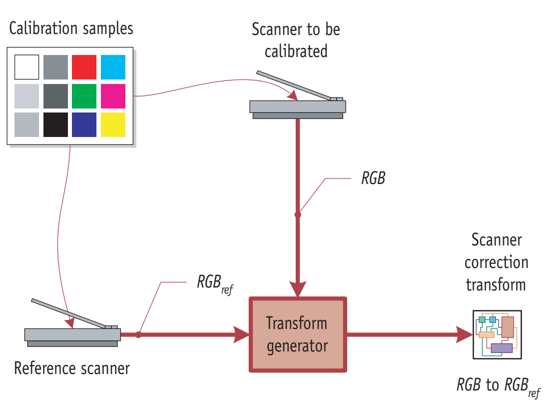


Figure 10.7a Scanner correction. Calibration samples are read on both the reference scanner and the scanner to be calibrated. A correction transform is then computed. This transform relates scanned *RGB* values to corresponding RGB_{ref} values for the reference scanner.

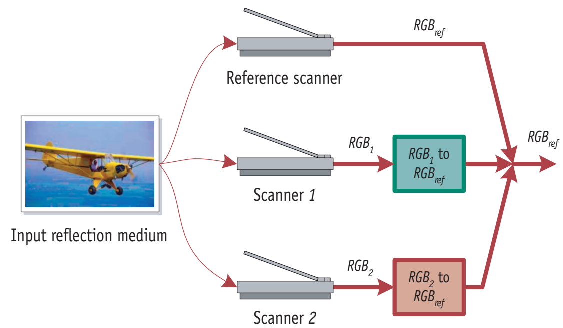


Figure 10.7b The use of multiple scanners and correction transforms. With appropriate correction transforms in place, each scanner produces *RGB* values equal to those that would be produced by the reference scanner.

When each scanner is used with its own correction transform, each will produce identical values for the same input image. If scanner characteristics change over time, the calibration can be updated at appropriate intervals to compensate.

It will be shown later that in color-imaging systems with disparate types of input media, input scanner calibration *alone* is not sufficient to create input compatibility. However, all successful color-imaging systems incorporate some form of scanner calibration-based correction as *part* of their input signal processing. Such corrections, which essentially make all system scanners behave identically, will be assumed in all further examples.

In one practical implementation of this calibrated densitometric approach, shown in Figure 10.8, all of the input scanners at a major motion picture special-effects studio were calibrated to produce *RGB* values equal to those measured according to ISO Status M densitometric standards. This simple calibration was *all* that was required to achieve input compatibility in this multi-million-dollar system because a single input medium—a particular motion picture color-negative film—always is used. Many electronic prepress systems also successfully use color encoding that is based solely on calibrated densitometric measurements. Scanner calibration is all that is required to create input compatibility in these systems, again because a single input medium (generally a particular photographic transparency film) is used.

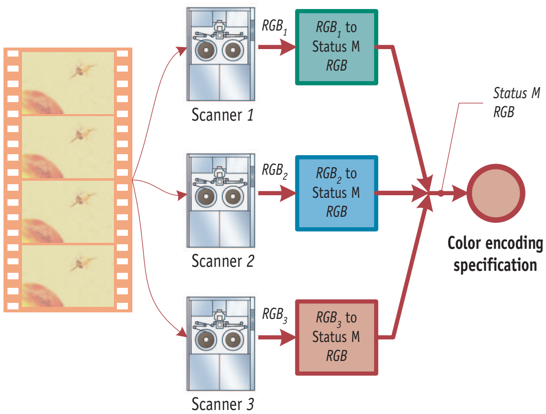


Figure 10.8 The use of densitometric color encoding in a motion picture special-effects studio.

But what happens if the system must support more than one input medium? Will densitometric color encoding still work? That depends on the nature of the differences among the media, as will be shown in later chapters. However, in the next example, color encoding based on densitometric measurement not only works for the described multiple input medium system, it essentially is the *only* type of encoding method that will work.

A multiple negative-film system

This next practical example is based on the Cineon digital film system (Figure 10.9), which has been used widely in the motion picture industry for special effects and other post-production work. The Cineon system is designed to scan multiple types of motion picture negative films for input to an electronic editing workstation. A film writer then is used to produce output on another type of motion picture negative film. The figure shows an example of three different negative films being scanned and edited together to form a single composite negative. Input compatibility is essential in this application in order to produce seamless-appearing composite images from portions of the three different input negative films.

How can input compatibility be created in this system? It was stated earlier that input compatibility must be based on some aspect of color that all of the system inputs have in common. In order to identify what the particular input films of this system have in common, it is necessary to know something about how they are designed to be used. The key piece of information is that these negative films are all intended to be printed onto a specific motion picture print film, using a particular type of optical printer. The negatives accordingly are designed in terms of a particular type of densitometric measurement described earlier: *printing density*.

As discussed in Chapter 8, the printing densities of a negative film are the densities of that film as measured according to a specified set of reference print material spectral sensitivities and a defined reference printer illuminant. The spectral characteristics of a particular reference print film and printer illuminant used for optically printing motion picture negatives are shown in Figure 10.10.

Recall that if two areas of a negative are "seen" as having the same *RGB* densities by the print material, the areas have—by definition—the same printing densities. If both areas have the same printing densities, they both will produce the same color when identically printed onto the reference print material. That will be true even if the areas having the same printing densities are on two different types of negative films having different sets of CMY image-forming dyes. Printing-density values defined according to a particular print medium and printer, then, are *unambiguous*. If the printing-density values for an area of a negative are known, no additional information as to the origin of those values is required. Color encoding based on printing densities therefore meets our criterion for input compatibility.

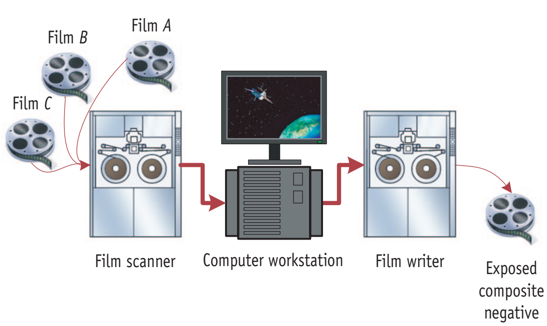


Figure 10.9 The Cineon digital film system.

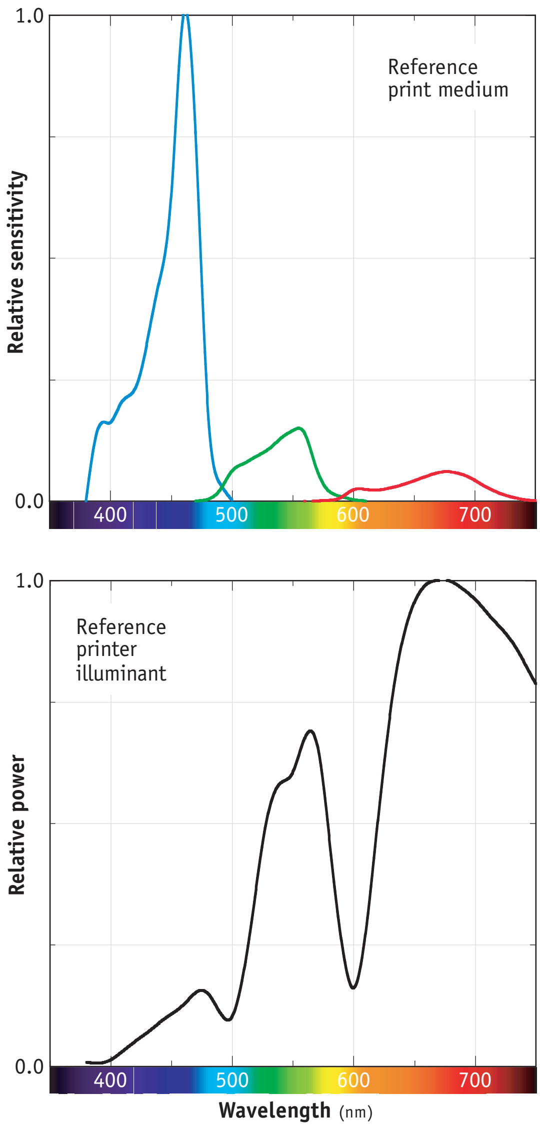


Figure 10.10 Spectral characteristics of a particular reference print film and printer illuminant used for motion picture negatives.

To encode scanned image data in terms of printing-density values, printing densities can be measured *directly* using a scanner having effective red, green, and blue spectral responsivities equivalent to the effective red, green, and blue spectral responsivities of the reference print material and printer. The effective spectral responsivities of a scanner (Figure 10.11a) are the product of the spectral power distribution of the scanner light source and the red, green, and blue spectral sensitivities of the scanner sensor. Similarly, the effective spectral responsivities for a print medium (Figure 10.11b) are the product of the spectral power distribution of the printer light source and the inherent red, green, and blue spectral sensitivities of the print medium.

If the effective spectral responsivities of the scanner are *not* equivalent to the effective spectral responsivities of the reference print material and printer, another approach to printing-density encoding can be taken. An input signal processing transform can be used to convert densitometric measurements from the scanner to printing-density values. In this case, however, the scanner would require a different input transform for each type of input film, i.e., for each film having a different set of image-forming dyes.

In the Cineon system, color encoding in terms of printing-density values not only results in seamless cutting and pasting, but also directly provides the information necessary to produce output negatives. These output negatives must be printing-density copies of the input negatives (unless deliberate editing has been done). The following example illustrates why.

In the original negative of a particular motion picture, an overhead microphone occasionally was visible during a scene. Reshooting the scene would have been prohibitively expensive. Instead, the involved frames were scanned on a Cineon system, and the resulting image data were digitally "retouched" to remove the microphone. The output negative frames then were physically spliced with the good frames of the original negative, and all of the frames then were optically printed onto the print film. None of this trickery was noticeable in the final print because, except for the deliberately retouched areas, the *printing densities* of the output negative perfectly matched those of the original negative. Any other type of matching would have produced unacceptable results because images printed from the retouched frames would have been noticeably different from those printed from the original frames.

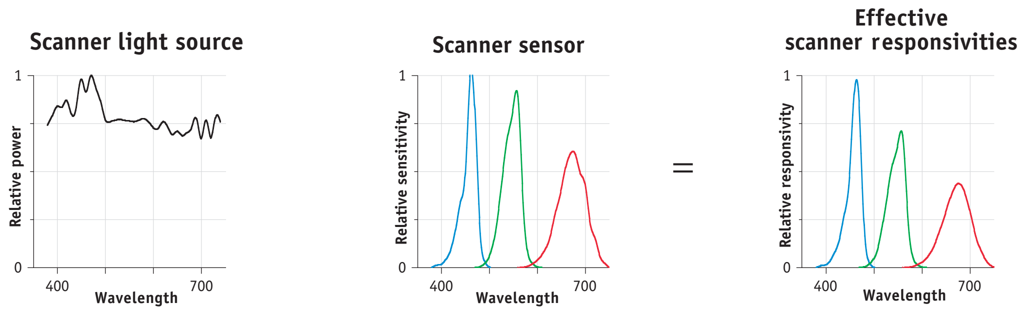


Figure 10.11a The effective spectral responsivities of a scanner are the product of the spectral power distribution of the scanner light source and the red, green, and blue sensitivities of the sensor.

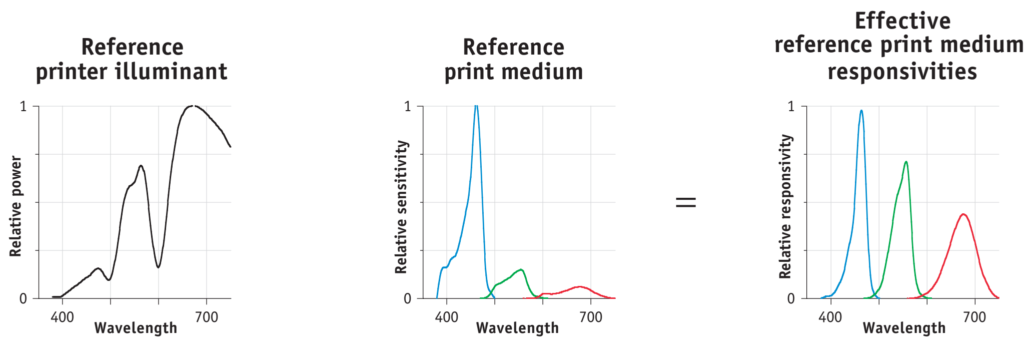


Figure 10.11b The effective spectral responsivities for a print medium are the product of the spectral power distribution of the printer light source and the red, green, and blue spectral sensitivities of the print medium.

Discussion

In the examples that have been described, the use of a densitometric color-encoding method resulted in encoded values that are *consistent* and *unambiguous* in their meaning. Those features are the essence of input compatibility, and they are the key to successful color encoding.

In some of the examples, other encoding methods could have been used successfully. But densitometric color encoding often is preferred because transformations from scanner *RGB* values to densitometric values, and transformations from densitometric values to output-device *RGB* values, generally are quite simple. That simplicity can translate to optimum signal processing accuracy and speed.

Although densitometric color encoding was successful in the examples discussed in this chapter, the input sources of the described systems were limited to particular types. In the following chapters, it will be shown that other encoding methods are needed in order to produce input compatibility in more complex, multiple input situations.

Summary of key issues

- Densitometric color encoding is based on color measurements made according to a defined set of spectral responsivities that are not equivalent to any set of visual color-matching functions.

- The principal advantage of densitometric color encoding is that it simplifies color signal processing, which can result in optimum signal processing speed and accuracy.
- Densitometric color encoding can be used only when an imaging system is restricted to having one basic type of input medium.
- All successful color-imaging systems incorporate some form of correction for input scanner variations. Such correction is based on calibration and is included in the input signal processing.
- Many color electronic prepress and motion picture systems successfully use color encoding based solely on calibrated densitometry.

11

Colorimetric Color Encoding

Colorimetric color encoding is based on measurements made according to the spectral responses of a human observer. One of the principal advantages of this method of color encoding is that it is based on well-established CIE recommendations for color measurement.

At first glance, colorimetric encoding would seem to offer the perfect method for encoding color, and, in practice, colorimetric encoding sometimes can provide input compatibility where methods based on other forms of measurements cannot. In other cases, however, the use of standard colorimetric techniques alone will not work.

Since colorimetric encoding is the method most often promoted as the ideal "device-independent" solution to color encoding, it is very important to understand its capabilities—and its limitations. The examples given in this chapter should help.

A case where colorimetric encoding works well

Consider a system, such as the one shown in Figure 11.1a, that supports input from an assortment of reflection media having different image-forming colorants—printing inks, photographic dyes, thermal-transfer dyes, etc.—with different spectral absorption characteristics. A color-encoding method based on red, green, and blue *densitometric* measurements alone, as was described in the previous chapter, will *not* provide a meaningful representation of color in this multiple reflection-media system.

For example, Figure 11.1b illustrates that a pair of colors on two different media can look identical, but they most likely will produce quite different *RGB* densitometric values. Conversely, a pair of colors on two different media might appear quite different from one another, but they might happen to produce the same *RGB* densitometric values. These inconsistencies will occur whenever the spectral absorption characteristics of the colorant sets used in the two media are different.

The meaning of measured *RGB* densitometric values alone, then, would be ambiguous in this system. The color associated with a given set of *RGB* values would depend on which particular medium was measured. Such ambiguity indicates that compatibility has not been established among the system inputs, and the system will fail unless something is done to establish that compatibility where it does not inherently exist.

It was stated earlier that in order to create a condition of input compatibility among disparate input media, it is necessary to identify a color-related property that all the involved media have in common. There is such a property in this example in that all the input media of the system were originally designed to be directly viewed by human observers. That fact suggests that the color encoding should be

Digital Color Management: Encoding Solutions, Second Edition Edward J. Giorgianni and Thomas E. Madden
 ISBN: 978-0-470-51244-9

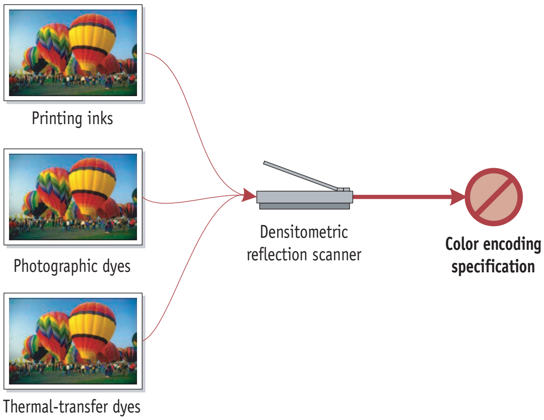


Figure 11.1a A multiple reflection-media system in which a densitometric scanner is used. Densitometric measurements alone will not provide a meaningful representation of color in this system.

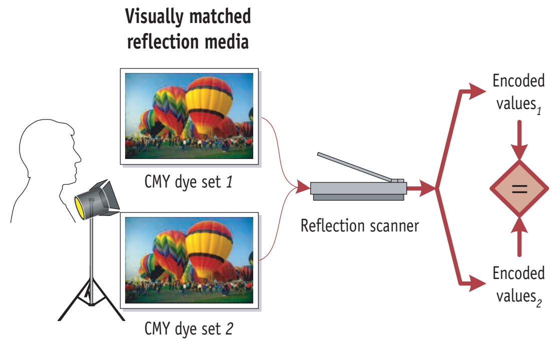


Figure 11.2 Input compatibility will be realized in this system if visually matched colors on the different input media produce identical encoded values.

based on *colorimetry* rather than on densitometry, and indeed that is true. However, that colorimetry must be measured in a very specific way.

Input compatibility will be realized, by definition, if visually matched colors on the different input media produce identical encoded color values (Figure 11.2). However, since the colorants differ from medium to medium, such visual matches will be metameric, not spectral. As discussed in Chapter 1, metameric matching is *viewing-illuminant dependent*; pairs of colors on different media that match when viewed under one illuminant might not match when viewed under another. What this means is that color encoding based on standard CIE colorimetric measurements *can* achieve input compatibility among multiple reflection media, but *only* under the following two conditions:

1. A reference illuminant, i.e., the illuminant used for metameric matching, must be specified.
2. Colorimetric values must be determined according to the spectral properties of that reference illuminant.

For example, suppose it is specified that two different input media will be encoded in terms of their being viewed under identical conditions in which a

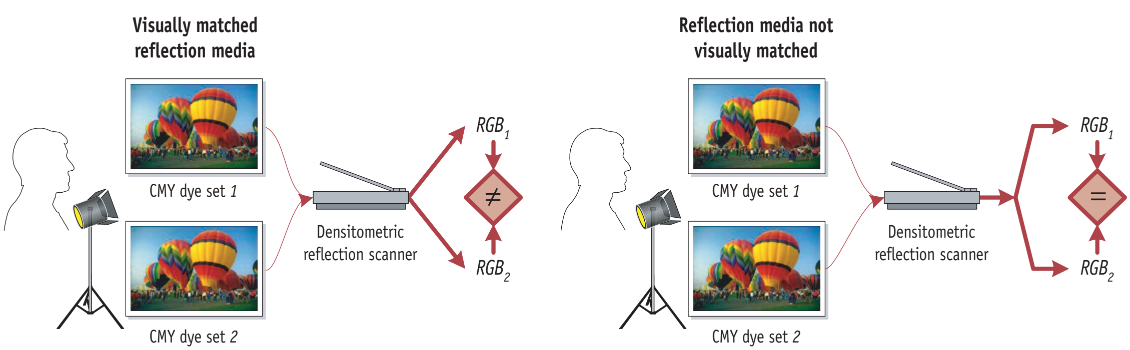


Figure 11.1b Color encoding based on red, green, and blue densitometric measurements alone will fail in this system. Densitometric color values are ambiguous when media having different colorants are involved.

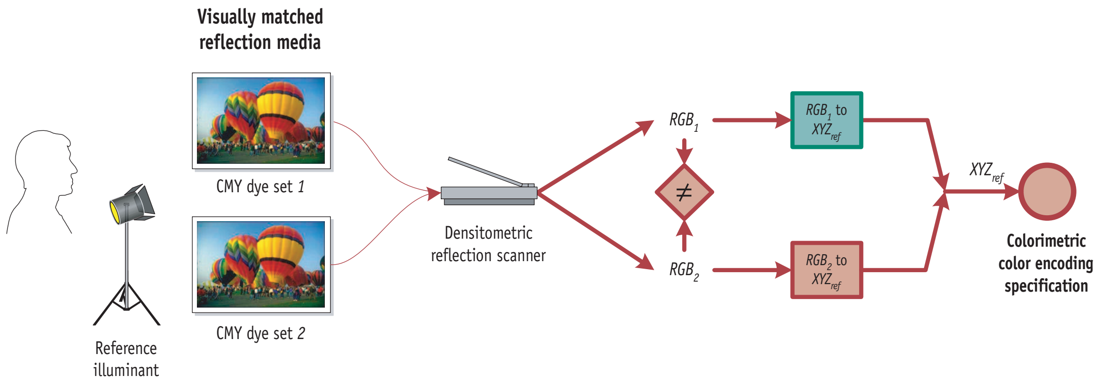


Figure 11.3 Media-dependent transforms are required when an RGB densitometric scanner is used for colorimetric color encoding.

light source that simulates CIE Standard Illuminant D_{50} is used. If a pair of colors on the two media visually match under those conditions, their *colorimetric* values measured using that D_{50} source would be identical. Color encoding derived from those identical values would properly represent the fact that the colors matched, which indicates that the colorimetric encoding method has successfully established input compatibility for the two media.

Colorimetric values of scanned input media can be determined in two fundamentally different ways. The first approach employs *densitometric* scanning (Figure 11.3). Media-dependent transforms are then used to convert *RGB* densitometric values to CIE colorimetric values that have been determined with regard to the reference illuminant. *It is important to note that a different transform is required for each input medium having a different set of image-forming colorants.* Implementing this approach, therefore, requires some means for identifying the input medium so that the proper conversion transform can be selected.

The second approach employs a *colorimetric scanner* to *directly* measure the colorimetry of the input media. Figure 11.4a shows one arrangement for a colorimetric scanner in which the spectral power distribution of the scanner light source is identical to that of the reference illuminant. In addition, the spectral sensitivities of the scanner sensors are equivalent to the color-matching functions of the CIE Standard Colorimetric Observer (or to any other set of visual color-matching functions). The values directly measured by this scanner are input compatible, so no media-dependent transforms are required.

The colorimetric scanner in Figure 11.4b is identical to that in Figure 11.4a except that the spectral power distribution of its light source differs from that of the reference illuminant. Due to this light source difference, colorimetric values measured by this scanner are *not* input compatible *with respect to the reference illuminant specified for the metameric matching*. For example, the scanner values for a pair of colors on two different media might be identical, but those colors might not visually match under the reference illuminant. Conversely, a pair of colors on two different media that visually match under the reference illuminant might produce different scanner values. To eliminate these inconsistencies, media-dependent input transforms once again must be used—in this case to convert actual measured colorimetric values to those that would have been measured under the reference illuminant. It is important to recognize that the need for these transforms largely negates the principal advantage of colorimetric scanning over densitometric scanning.

The colorimetric scanner shown in Figure 11.4c again uses a light source having a spectral power distribution different from that of the reference

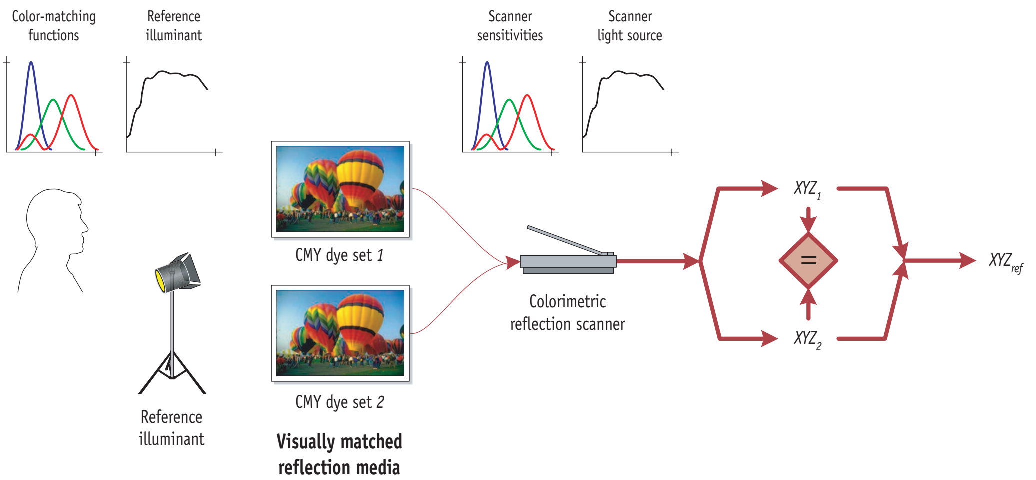


Figure 11.4a Use of a colorimetric scanner for colorimetric color encoding. In this arrangement, the spectral power distribution of the scanner light source is identical to that of the reference viewing illuminant. Measured values are input compatible, so no media-dependent input transforms are required.

illuminant. In this case, however, the spectral sensitivities of the scanner sensors are *not* equivalent to a set of visual color-matching functions. Instead, the sensitivities are designed such that the *effective spectral responsivities* of the scanner, i.e., the products of the spectral power distribution of the scanner light source and the spectral sensitivities of the scanner sensors, are equivalent to the product of the spectral power distribution of the reference illuminant and a set of visual color-matching functions.

In the example shown in Figure 11.4c, the spectral sensitivities of the sensors have been designed

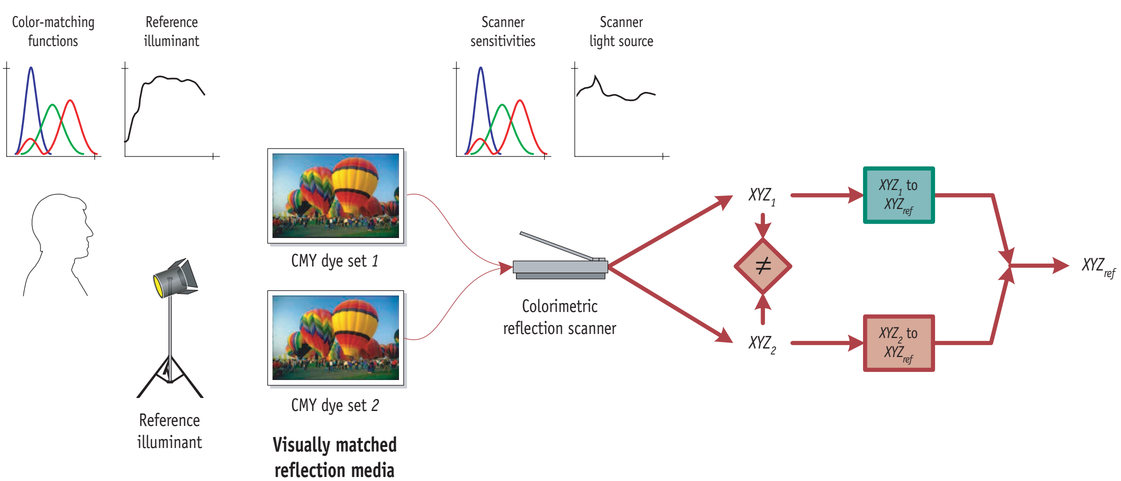


Figure 11.4b Media-dependent input transforms are required in this colorimetric scanner because the spectral power distribution of its light source differs from that of the viewing illuminant specified for metameric matching.

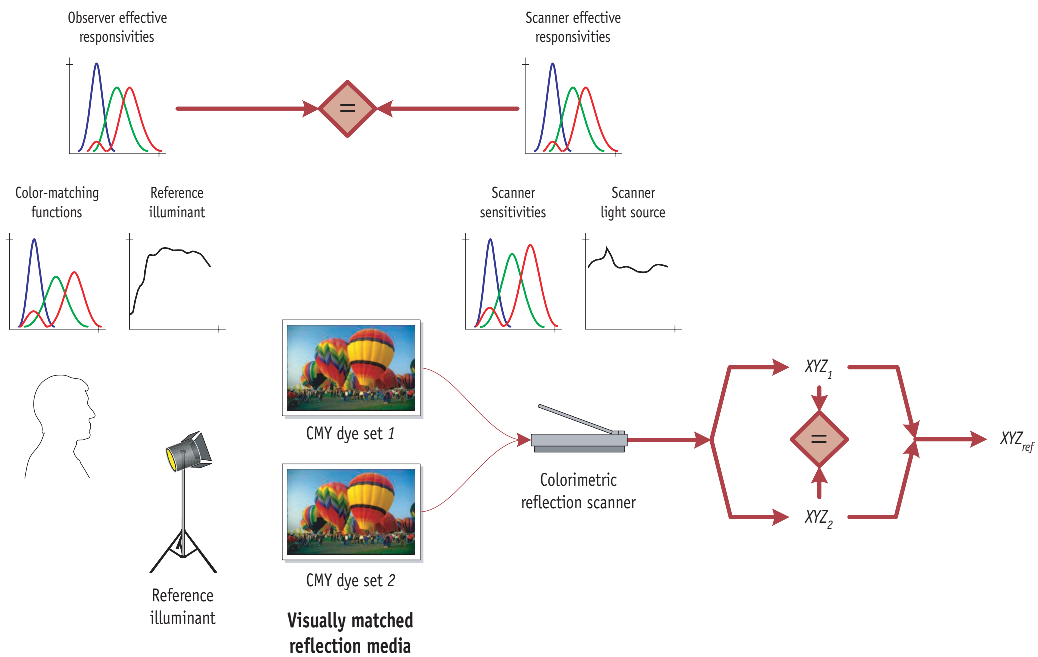


Figure 11.4c In this colorimetric scanner, the effective spectral responsivities are equivalent to those of the scanner in Figure 11.4a. Measured color values are input compatible, thus media-dependent transforms are not required.

such that the effective spectral responsivities of the scanner are identical to those of the scanner described in Figure 11.4a, despite the fact that a different scanner illuminant has been used. Except in some special circumstances, generally involving input media having fluorescent properties, the color values measured by the two scanners will be identical. Since those measured values are input compatible, no media-dependent input signal processing transforms are required.

A case where colorimetric encoding provides a partial solution

Colorimetric color encoding can work well in systems that are limited to reflection-media input because reflection images typically are viewed in environments containing white objects and other visual cues that strongly influence the general-brightness adaptive state and chromatic adaptive state of the observer. As a result, there generally is good correlation between color appearance and measured colorimetry. For example, a reflection print that appears too dark or too green in color balance also will measure that way. That is not the case for all image types, however. In particular, it generally is not the case for projected images such as those produced by photographic slide films, motion picture print films, and electronic projectors.

Figure 11.5a, for example, shows an example system with input from two different photographic slide films. Since these films both are designed to be viewed directly by a human observer, it might seem logical to encode the inputs according to CIE colorimetric measurements. As in the previous example, this could be accomplished either by the use of a colorimetric scanner or by the use of a densitometric scanner and appropriate media-specific transforms. In this situation, these standard colorimetric approaches will achieve input

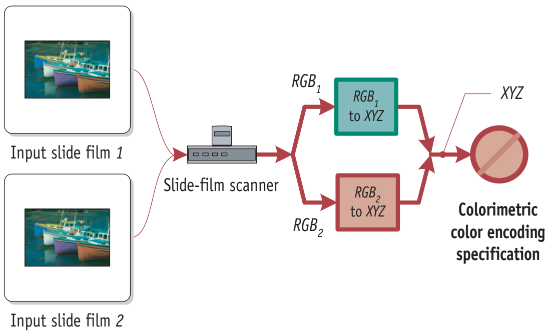


Figure 11.5a A system using multiple types of photographic slide films. Standard colorimetric encoding will achieve input compatibility in this example, to a degree. But there is a complication that may not be obvious, as explained in the text.

compatibility, *to a degree*. But there is a complication, which may not be obvious, that applies not only to images from slide media but to all forms of hardcopy and softcopy images intended for viewing in darkened environments.

The problem is that when an image is viewed in an otherwise dark viewing environment, there can be a significant amount of visual adaptation by the observer to the image itself. As discussed in Chapter 7, this adaptation to the image occurs because other visual cues and references are absent. As a consequence, fairly large shifts in image colorimetric characteristics may not be obvious to an observer.

This can be demonstrated by sequentially projecting, in a darkened environment, a series of images of the same subject matter in which the color balance is slightly bluer in each successive image. If given a few seconds to adapt to each image, most observers do not detect any image-to-image differences even though the colorimetry of the final image indicates it is much bluer than the first. This demonstration and a similar one described in Chapter 7, in which the overall luminance level of each projected image was decreased, show that the relationship between *colorimetric measurement* and *color appearance* is not straightforward for images viewed in conditions where there is significant adaptation to the image itself.

That is a problem when images intended for such viewing conditions are used for input to color-imaging systems. The colorimetry of these images can be expected to vary because of original-scene illuminant differences, exposure errors, media or device manufacturing variability, chemical or electronic signal processing variations, and a host of other factors. Although they might not be noticeable to an observer viewing projections of the images, those variations will be faithfully measured and encoded when standard colorimetric encoding methods are used.

Whether the resulting encoded differences will be visually noticeable in subsequent output images will depend on what type of output viewing environment is used. If the output is another dark-surround display, any image-to-image variations will be no more visually noticeable than those of the input images. If, however, the output is to a different type of display, such as a reflection print viewed in a normal environment, visually undetected input-image colorimetric variations may be unacceptably noticeable in the output images. Image-to-image variations also might be very apparent, on *any* form of output, in a single composite image made from portions of different input images or in a display in which multiple images are viewed simultaneously.

This is a tough problem to solve. What generally is done in practice is that input images are first previewed on a softcopy display (Figure 11.5b), and the system operator adjusts data from each input image for overall lightness and color balance before the final encoding is performed. This process requires considerable skill and patience. It will be shown later how a scene-based color-encoding method greatly facilitates the use of algorithms that can perform this operation automatically.

A case where colorimetric encoding does not work

Figure 11.6 shows a system supporting representative examples of *both* input types that have been discussed in this chapter—a photographic slide film and a reflection-print medium. Since both media are designed to be viewed directly by human observers, it might seem that encoding according to CIE

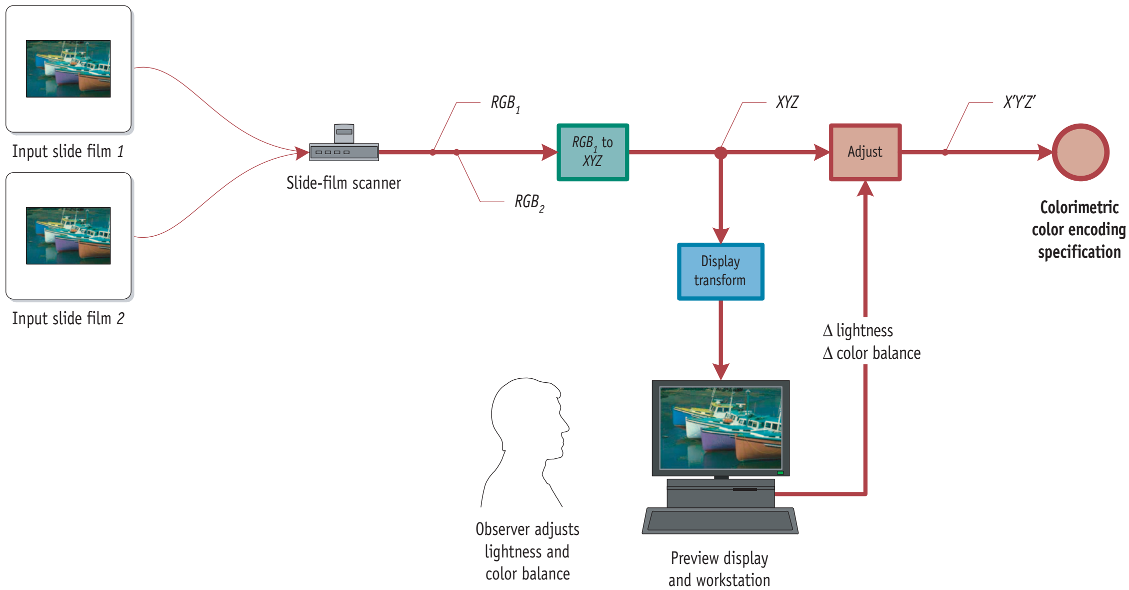


Figure 11.5b Adjusting scanned image data prior to encoding.

colorimetric measurements should work at least reasonably well. But standard colorimetric encoding would in fact drastically fail to achieve input compatibility for this combination of input types. The reason is that the example input media are designed to produce images intended to be viewed in very different environments. Their colorimetric characteristics therefore are fundamentally different and inherently incompatible.

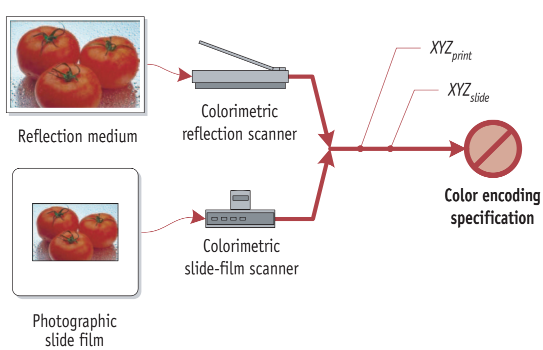


Figure 11.6 A system supporting both a photographic slide film and a reflection medium. The use of standard colorimetry alone will not work.

As described in Chapter 7, three types of colorimetric adjustments that are *not* required for reflection media are built into photographic slide films. First, the overall image density is increased in order to allow for an extension of the exposure dynamic range. This can be done on a slide film because the overall higher density goes unnoticed due to the observer's general-brightness adaptation to images viewed in a dark-projection environment. Second, the photographic gamma of the grayscale is increased to compensate for the reduced luminance contrast an observer will perceive due to the dark areas surrounding a projected image. And third, the color balance is shifted cyan-blue to compensate for the observer's incomplete chromatic adaptation to the reddish-yellow color of the projector light source.

Obviously, a measuring device such as a colorimetric input scanner is not subject to these perceptual effects, so its measurements will not necessarily correspond to actual color appearances. A colorimetric scanner therefore would measure a photographic

Figure 11.7 A normal reflection print (left). A reflection print made according to colorimetry measured from a photographic slide (right) will be too dark, too high in luminance contrast, and too cyan-blue in color balance.

slide as being darker, higher in luminance contrast, and more cyan-blue in color balance than the projected slide actually appears. Those characteristics would be readily apparent in an output reflection image made according to the slide's measured colorimetry (Figure 11.7).

In this dual-medium example system, then, the meaning of a set of standard colorimetric values would be ambiguous. The color appearance associated with those values would be entirely different depending on whether the values were produced by scanning a slide film or by scanning a reflection medium. In this system, then, colorimetric encoding will *not* consistently yield a proper interpretation of the meaning of the input images. As a result, a state of input compatibility will not be achieved. Unless other measures are taken, the system will fail.

This outcome is not limited to isolated cases. A color-encoding method based on standard CIE colorimetric measurements *alone* will fail any time it is used for a mix of input-image types that are designed to be viewed in different environments. But if standard colorimetric encoding fails, what else can be done?

There are two basic alternatives, each having its own advantages and disadvantages. One is to encode in terms of the *reproduced color appearance*, rather than the measured colorimetry, of the input image. This involves the use of advanced forms of colorimetry that appropriately account for perceptual factors. That approach will be described later, in Part IV, as part of the discussion of color management. Another approach, which the authors originally devised for the Kodak *Photo CD* System, will be described in the following chapter.

Discussion

The rigor and elegance of CIE colorimetric methods can lead the unwary to believe that color can be readily encoded by straightforward applications of standardized physical measurements and computational techniques. However, the examples in this chapter have shown that standard CIE colorimetric techniques alone are not always sufficient for meaningful color encoding. Lack of understanding of this point has been responsible for the failure of numerous color-imaging systems, color-management products, and color-encoding methods.

This should not be interpreted as a criticism of CIE colorimetry. The CIE system was designed for particular types of applications—primarily those involving the color matching of samples under identical viewing conditions—that have little to do with imaging. For more than 75 years, the system has been shown to work extremely well for such applications. What is being discussed here is the inappropriate usage of the CIE system.

One should be wary of systems that seem to "prove" the applicability of colorimetric color encoding. In most cases, such systems are limited to reflection input and output media. Since there generally is good correlation between standard colorimetry and color appearance for reflection media, these systems will work reasonably well. When other types of media are used, however, the systems fail. Some systems that work across media types (such as some prepress systems that use photographic transparency film input and reflection-print output) may claim to be using standard CIE colorimetry. If these systems are examined carefully, however, it will be found that input colorimetric values are significantly altered before being sent to the output device. Somewhere in the signal processing path, colorimetric values measured from the transparency grayscale are transformed to those of a typical reflection print. Without the use of such colorimetric alterations, these "colorimetric" systems do not work.

Unfortunately, as such limitations have become more widely acknowledged, a certain backlash has resulted. It is now not unusual to hear it stated, in effect, that CIE colorimetry does not work, so it should be abandoned. Abandoning CIE colorimetry would be at least as big an error as assuming that it is entirely sufficient for all color-encoding situations. As will be shown in upcoming chapters, colorimetry can still form the basis for encoding methods that achieve true input compatibility in complex color-imaging systems.

Summary of key issues

- Colorimetric encoding is based on measurements made according to the spectral responses of a standard human observer.
- Standard colorimetric encoding works well only in situations where there is good correlation between colorimetry and color appearance for the system media.
- Colorimetric encoding requires that the conditions for input-media metamerism be specified.
- A true colorimetric scanner has effective spectral responsivities equivalent to the product of a set of visual color-matching functions and the spectral power distribution of the viewing illuminant specified for input-media metamerism.
- An RGB densitometric scanner that uses a transform to generate colorimetric values is not a true colorimetric scanner. Media-specific input transforms are required in such scanners.
- Standard colorimetric encoding alone is insufficient for achieving input compatibility when there is inconsistent correlation between colorimetry and color appearance for the system media.

12

Scene-Based Color Encoding

In 1992, the Kodak Photo CD system (Figure 12.1) was introduced by Eastman Kodak Company. This system, which was quite sophisticated for its time, produced digital images from photographic negatives, transparencies, and reflection prints and stored them on CDs. Digital images from other sources such as electronic cameras also could be stored on the same CDs. Photo CD discs could be used to produce video images on television receivers and monitors using special disc players. Images of various resolutions also could be displayed on computer systems using CD-ROM drives and appropriate software. Hardcopy output images, such as 35 mm slides, large-format transparencies, reflection prints, and color separations for graphic arts applications, could be produced from Photo CD discss on systems equipped with film writers, thermal printers, and other types of digital hardcopy output devices.

Our main assignment in the development of this system was to devise a method for digitally encoding color on Photo CD discs. Because the system was specified to include a disparate assortment of inputs and both hardcopy and softcopy outputs, the task turned out to be an interesting challenge. Ultimately, it led to our invention of an entirely new method of encoding color, which we refer to as *scene-based color encoding*. That method is now an important component of advanced color-imaging systems such as the one described in detail in Chapter 21 currently being designed for digital cinema. This chapter describes the function and methodology of scene-based color encoding by way of a first-hand account of its development for the Photo CD system.

Encoding approaches

We began our work on the Photo CD system by considering the applicability of existing methods for encoding color. Various encoding methods based on RGB densitometric or standard CIE colorimetric measurements alone were ruled out immediately, since neither approach would have worked for the disparate mix of input types specified for the system.

We considered and also ruled out encoding schemes that would have been entirely dependent on the use of image file identification tags. Photo CD format image files do indeed include tags, but the tagged information is supplemental. As was discussed earlier, a system that is entirely dependent on tags requires some form of intermediary signal processing such as that provided by a workstation equipped with an appropriate color-management application. In the absence of that intermediary processing, every output device would be required to have a complete array of output transforms—one for every possible type of input. That requirement would have made Photo CD players and similar devices far too expensive and complex. Moreover, the required amount of output signal processing would

Digital Color Management: Encoding Solutions, Second Edition Edward J. Giorgianni and Thomas E. Madden
 ISBN: 978-0-470-51244-9

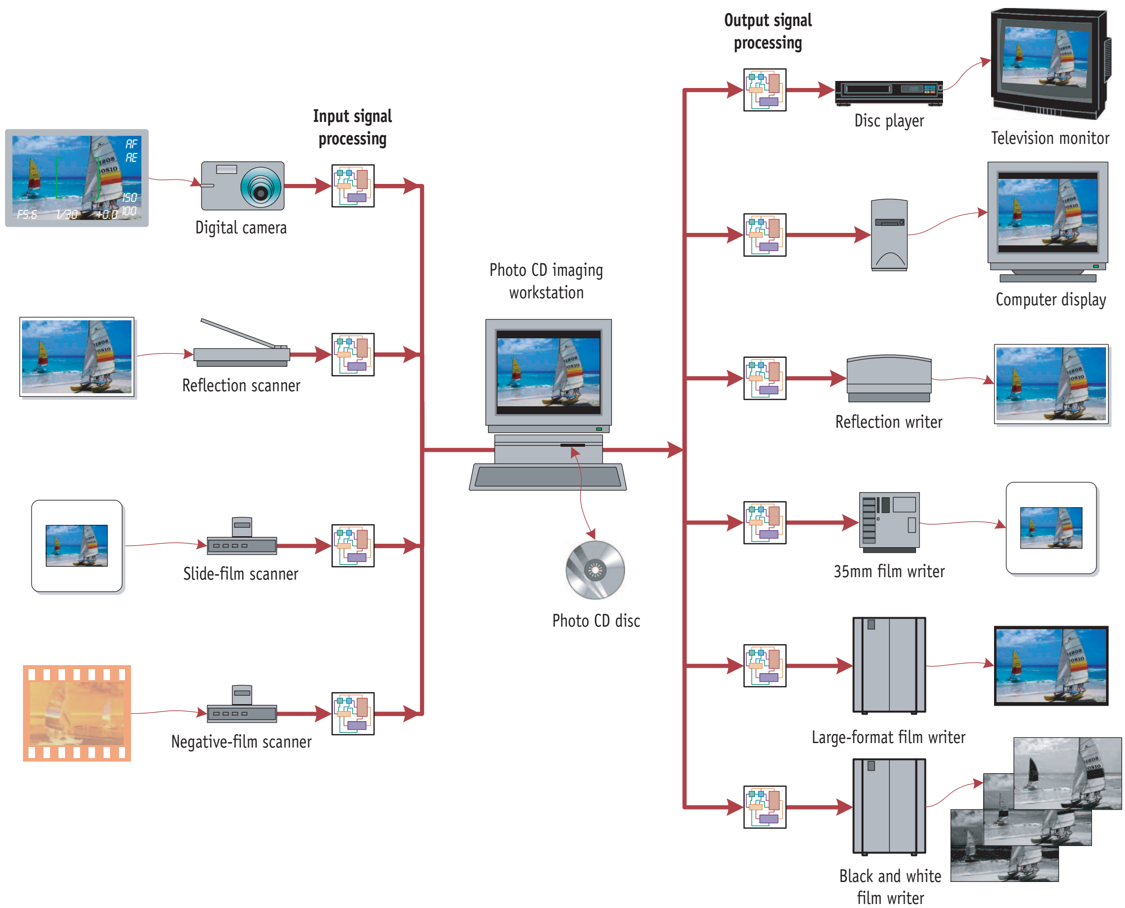


Figure 12.1 The Kodak Photo CD system. The system supported a disparate array of input sources and output devices.

have prevented rapid video displays of Photo CD format images on most desktop computer systems.

We gave very serious consideration to using some type of appearance colorimetry to encode images in terms of their reproduced color appearance. As briefly described earlier, appearance colorimetry begins with standard colorimetric methods, but it includes adjustments for various perceptual effects. In addition to its complexity, however, this method had two drawbacks.

First, for input compatibility, a color negative must be encoded not in terms of its own appearance but in terms of the appearance of a print made from that negative. This can be accomplished by appropriate transformations of data scanned from the negative. The principal problem with the approach is that some information that was recorded by the negative is irretrievably lost in this type of encoding (Figure 12.2). This fact was particularly important for the Photo CD system. As the system was originally envisioned, negatives were to be the principal form of input. It would not have made much sense to design a system that compromised the quality of image data derived from the system's principal source of input.

Moreover, the system design specifications included software tools that would allow shifts in color balance and lightness to be applied to images *after* they had been encoded and recorded on disc. Such shifts, when applied to images from negatives, should

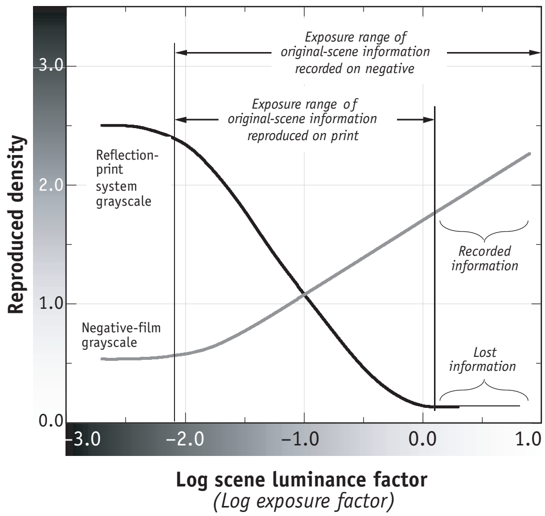


Figure 12.2 Printing a photographic negative results in the loss of some recorded information. Encoding in terms of print color appearance can result in a similar loss of information from the negative.

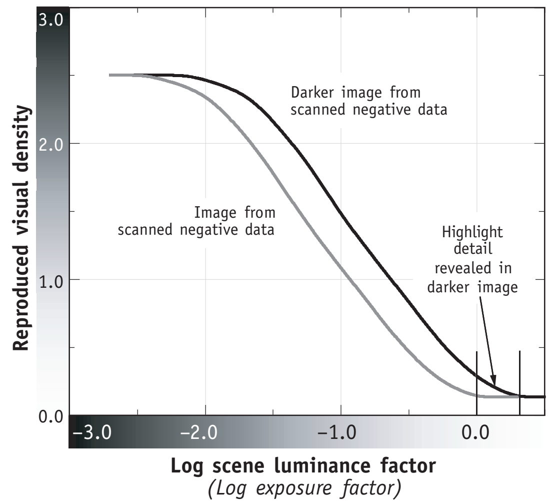


Figure 12.3a Producing a darker image from data scanned from a photographic negative should reveal more information in highlight areas, as shown in the grayscale characteristic for the darker image. The greater slope of the grayscale in the marked region indicates that additional scene highlight information will be visible in the output image.

result in new tonal information being made visible. For example, if image data encoded directly from a negative are shifted to produce a lighter output image, details that were not previously visible in dark shadow areas of that image should be revealed. If the encoded data are shifted to produce a darker output image, details that could not be seen previously in highlight areas likewise should be revealed (Figure 12.3a).

However, making similar shifts using data encoded in terms of the color appearance of a *print* made from a negative would not produce comparable results. For example, if such data are shifted to produce a darker output image, the highlight areas of that image simply would become darker. No additional highlight information would be revealed because it is not present in the encoded print data (Figure 12.3b). In Part IV, color-appearance encoding techniques that minimize the loss of information from color-negative films and other types of high-dynamic-range inputs will be described. But for the Photo CD system, the complexity of the overall scheme would have made it impractical.

The second problem with encoding based on reproduced color appearance relates to a topic

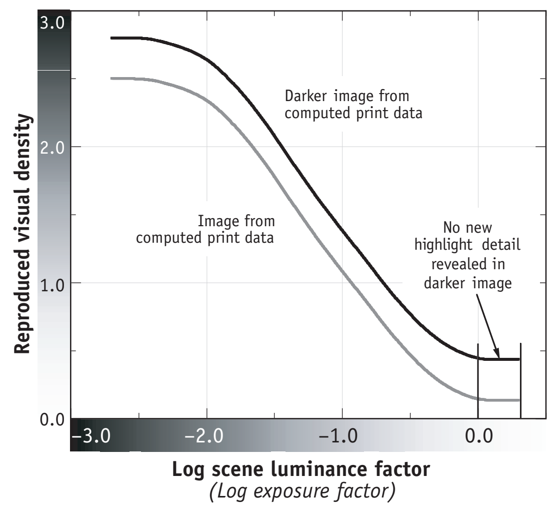


Figure 12.3b A darker image produced from corresponding print data will not contain any additional information. The slope of the grayscale characteristic for the darker image is unchanged in the marked region, which indicates that no additional scene highlight information will be visible in the output image.

discussed earlier: each type of input medium, and each individual medium within the different types, has its own distinctive "look" or "personality." For example, some media are considerably higher or lower than others in attributes such as grayscale contrast, luminance dynamic range, and overall color saturation. Although such differences might not always be apparent when images on different media are viewed separately, they become *very* noticeable when images are viewed simultaneously or are merged to form composite images. This was an important concern because it was expected that Photo CD discs would be used in applications such as desktop publishing and scrapbooking to create composite images and to produce individual pages containing multiple images.

A related concern was that the differences among images from various input types would be noticeable when Photo CD format images were viewed in succession. That was important, because it was felt that Photo CD discs frequently would be used to display sequences of images, as is done in slide presentations. Tests proved that this concern was indeed justified. It was found that if discs were created containing images from a variety of input media, and the appearance of each input image was preserved, the input media differences were very apparent when disc images where shown in sequence. The overall effect was quite distracting. Images seemed to "jump" whenever the original input medium changed from one type to another. As a result of those tests, producing *consistency* in the appearance of output images became an important objective.

The complete list of objectives for the color-encoding method then was as follows:

- To support input of images from photographic negatives, transparencies, reflection prints, and electronic cameras.
- To preserve the extensive range of original-scene information recorded on both positive and negative photographic media.
- To provide for the production of consistent, high-quality images on all supported output media and devices.

Since no existing color-encoding method met all of these objectives, we needed to invent and develop a new one, which we refer to as scene-based color encoding.

A new approach for the Photo CD system

The basic approach that we took can best be explained by returning to the multiple language text-scanner analogy that first was discussed in Chapter 9 (and shown again in Figure 12.4). Recall that

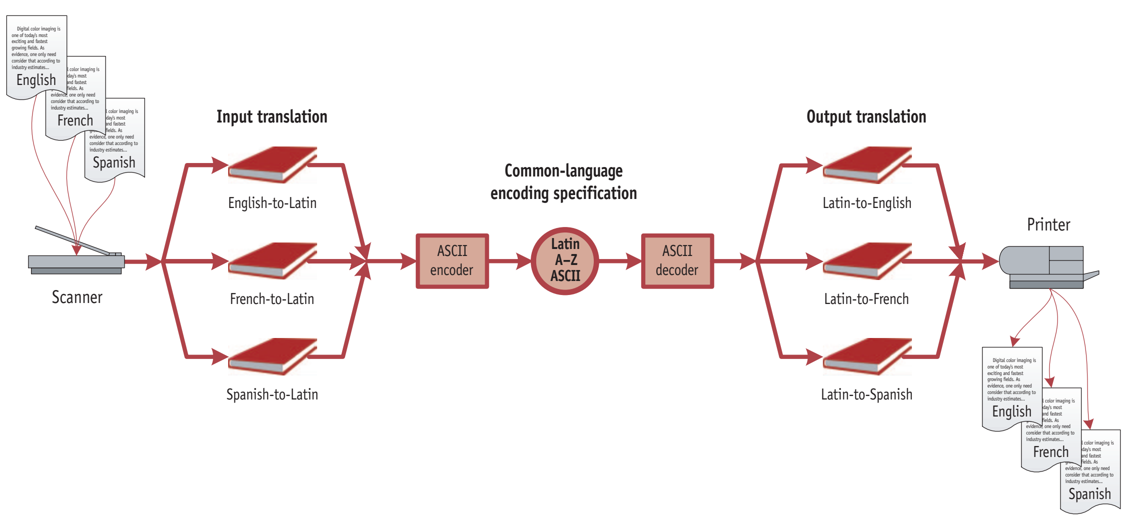


Figure 12.4 A multiple language text-scanning system.

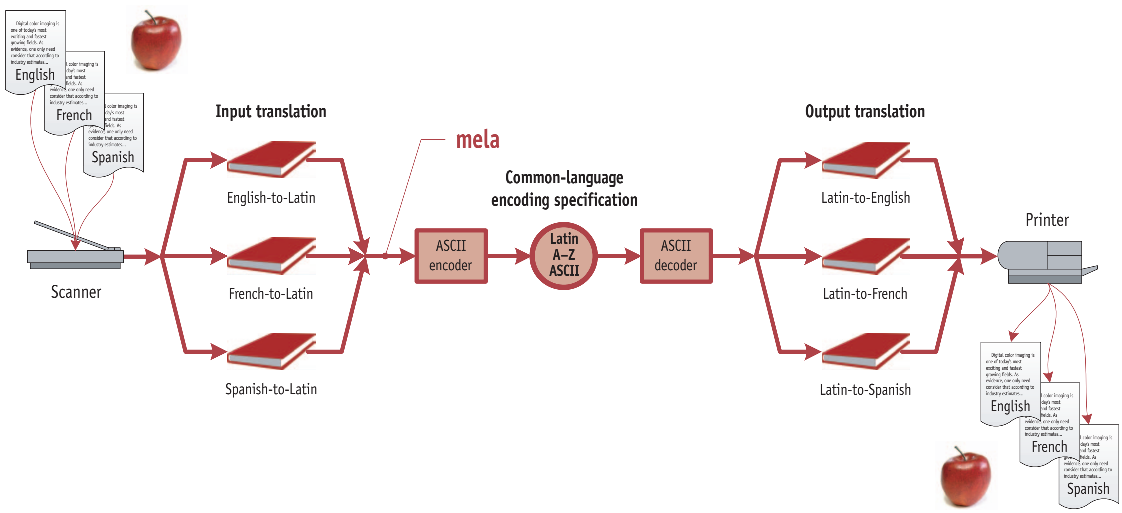


Figure 12.5 The basis for input compatibility in this text-scanner system is an inherent commonality that is independent of the languages involved. “So that which we call an apple by any other name . . .”

input compatibility was achieved in that system by the translation of words from each input language to a common encoding language. It is important at this point to understand exactly *why* that approach worked.

The reason is that there is a basic *commonality* among all languages. Although different words are used in different languages, they all represent objects, ideas, etc., that exist independently of the words used to describe them. For example, consider the apple shown in Figure 12.5. Regardless of what word or what language is used to *name* or “encode” that apple, it is still the same apple. In the figure, it is represented by the Latin word for apple, *mela*, but any language could have been used. The basis for input compatibility in the text-scanner system, then, is an inherent commonality that is independent of the languages involved.

Figure 12.6 shows that in an analogous imaging situation, the same inherent commonality exists. If an apple were photographed using three different types of media and the resulting images were scanned, very different measured values would result. In effect, there would be three different “translations” of the same original object color. The signal values from the digital camera would represent yet a fourth “translation.” But if it could be determined exactly *how* the apple’s color has been “translated” to color values by each input, it should be possible to “translate” *backwards* from those values to recover the fact that the same apple color was indeed photographed by each medium.

This line of thinking suggested an entirely different approach to color encoding. Instead of encoding in terms of color *translations* represented by the *reproduced* colors of each input, it should be

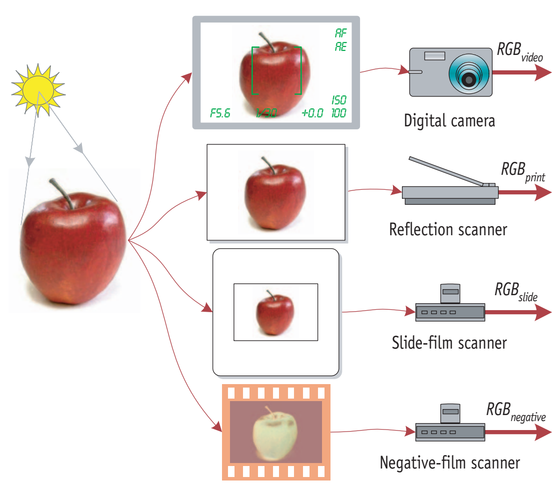


Figure 12.6 These four sets of *RGB* color values represent different “translations” of the apple’s color.

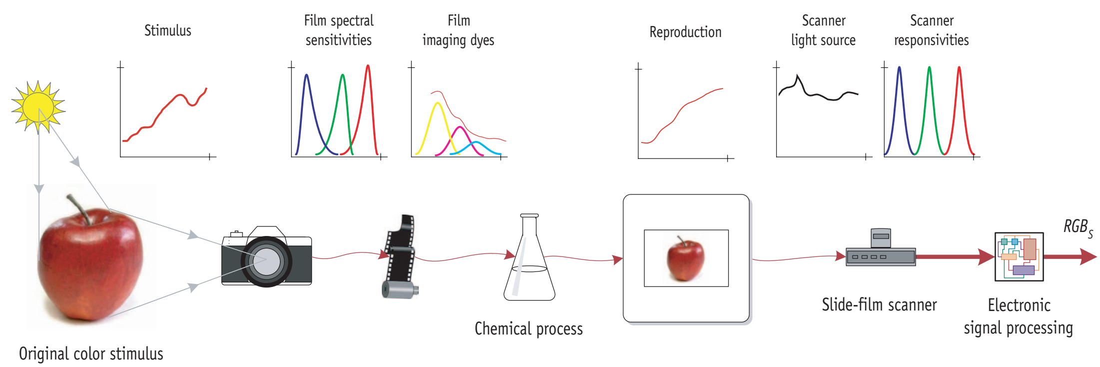


Figure 12.7 Translation of a color stimulus to scanner *RGB* values R_S, G_S, and B_S.

possible to encode in terms of *interpretations* of those translations. In other words, it should be possible to encode in terms of *original-scene colors*.

It was clear that this new scene-based color-encoding approach, if feasible, would accomplish all of the objectives of the system:

- The color encoding would support input from *any* source, because the fundamental incompatibilities among the inputs would be removed by the process that "translates" back to original-scene colors.
- Because it is not based on reproduced colors, the approach would allow the preservation of all original-scene information recorded by the input device or medium.
- The approach would allow the production of consistent output images, because the encoding itself would be consistent. It would be free of the color-reproduction differences that exist among the input sources.

In theory, the approach seemed ideal for the Photo CD system. But what about its practical implementation? Would it really be possible to interpret original-scene colors from reproduced "translations" of those colors?

Making the scene-based approach work

In order to investigate the possibility of interpreting original-scene colors from hardcopy reproductions, we first needed to take a detailed look at the translation process from original-scene colors to reproduced color values, as measured by an input scanner. That process is illustrated, for an original color stimulus and its reproduction on a photographic transparency film, in Figure 12.7.

As the figure shows, the stimulus is captured according to the film's red, green, and blue spectral sensitivities. The resulting exposure signals (recorded in the latent image) later are chemically processed, and an image is formed by the cyan, magenta, and yellow dyes of the film. The end result is a color area on the film, having a particular spectral transmittance. When that area is scanned, scanner *RGB* values R_s, G_s, and B_s are produced. These values depend on the spectral transmittance of the color area, the scanner illuminant, the red, green, and blue spectral responsivities of the scanner, and any electronic signal processing that might operate on the scanned signals.

Because of the chemical interlayer interactions and other complexities of modern photographic films, the relationship between the original color stimulus and scanner *RGB* values is not at all straightforward. Two approaches were employed in attempting to quantify that relationship.

The first involved mathematically emulating the color-reproduction characteristics of the film and scanner, using computational models previously developed for designing and optimizing new photographic products. The second approach treated the input film and the scanner as a single "black box." The properties of this black box could be characterized, from experimental data, by photographing a variety of original color stimuli and scanning the resulting color patches produced on the film.

The strength of a black-box characterization technique is that it can provide very accurate data, since it is based on actual measurements. However, the results apply only to the specific sample of film and the particular scanner used for the measurements. Mathematical models generally are not quite as accurate, but their results are free from the normal variabilities of film manufacturing, chemical processing, scanner calibration errors, and so on.

In the end, the best description of the relationship between original color stimuli and scanner values was obtained from a combination of the two approaches. The combined approach is based on characterization, but modeling is included to account for any differences between the actual film and chemical process and a nominal film and process. The nominal film and process are based on manufacturing specifications for the particular type of film being used. Compensations also are included for any differences between the actual scanner and a nominal scanner, where the nominal scanner also is based on manufacturing specifications.

Once the relationship between original stimuli and scanner values was determined, the inverse of that relationship could be used to convert scanner values to original stimuli values. This required a method for representing the original color stimuli in terms of three data channels. For that purpose, and others that will be described later, we created a mathematical concept we called the Photo CD reference image-capture device.

Figure 12.8 shows this hypothetical reference image-capture device and an individual color stimulus. When the device captures light from the color stimulus, it produces *RGB* exposure-factor values. More will be said in the next chapter about the specific characteristics of the reference device, but it is important to note here that its spectral responsivities correspond to a set of color-matching functions. The use of such responsivities is critical, because it means that *the RGB exposure-factor values of the reference image-capture device are tristimulus values that correspond directly to CIE tristimulus values of original color stimuli.*

Figure 12.9 shows the application of the basic encoding approach to a single input film. A number

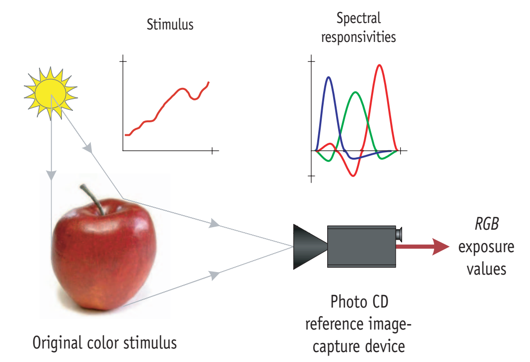


Figure 12.8 A reference image-capture device and a color stimulus. The *RGB* exposure-factor values of the device are tristimulus values that correspond directly to CIE tristimulus values of original color stimuli.

of original color stimuli, with known spectral power distributions, are photographed. In most cases, the stimuli would consist of a light source of known spectral power distribution and color patches of known spectral reflectances. The exposed film is chemically processed and scanned. The scanned values are corrected for any deviations of the actual film, chemical process, and scanner from a nominal film, process, and scanner to form *RGB* values R_s, G_s, and B_s.

Reference image-capture device exposure-factor values are then calculated for each original color stimulus, using the following equations:

$$\begin{aligned} R_{exp} &= k_r \sum_{\lambda} S(\lambda)\, R(\lambda)\, \bar{r}(\lambda) \\ G_{exp} &= k_g \sum_{\lambda} S(\lambda)\, R(\lambda)\, \bar{g}(\lambda) \qquad (12.1) \\ B_{exp} &= k_b \sum_{\lambda} S(\lambda)\, R(\lambda)\, \bar{b}(\lambda) \end{aligned}$$

where R_{exp}, G_{exp}, and B_{exp} are the device's red, green, and blue exposure-factor values; $S(\lambda)$ is the spectral power distribution of the light source; $R(\lambda)$ is the spectral reflectance of a color patch; $\bar{r}(\lambda)$, $\bar{g}(\lambda)$, and $\bar{b}(\lambda)$ are the red, green, and blue spectral responsivities of the device; and k_r, k_g, and k_b are normalizing factors determined such that R_{exp}, G_{exp}, and $B_{exp} = 1.0$ when the stimulus is a perfect white. Finally, a mathematical transform is built relating the scanner

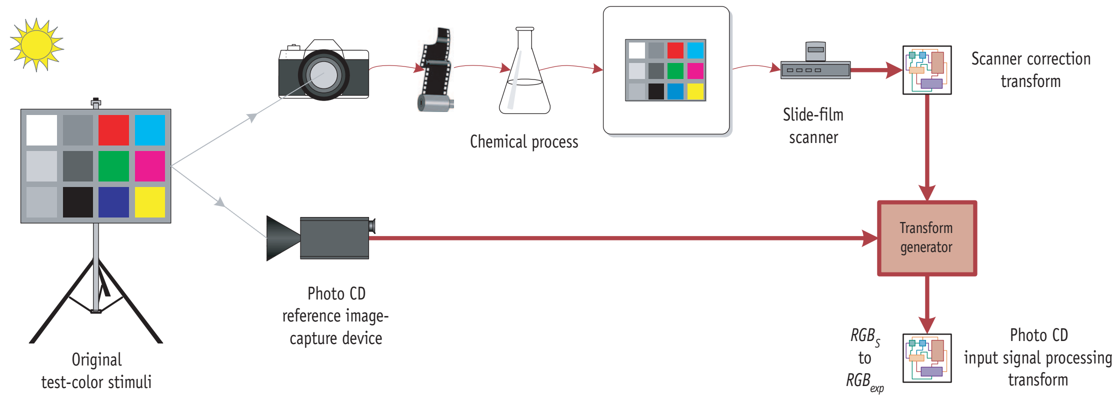


Figure 12.9 Building a Photo CD system input signal processing transform. The transform relates scanner *RGB* values R_s, G_s, and B_s to reference image-capture device exposure-factor values R_{exp}, G_{exp}, and B_{exp}.

R_s, G_s, and B_s values to the device R_{exp}, G_{exp}, and B_{exp} values.

Figure 12.10, which shows the relationship of the reference image-capture device to actual Photo CD system inputs, illustrates the overall concept of scene-based color-encoding. Color-image signals from a reference image-capture device (if one actually existed) would be encoded directly, without additional input signal processing. Signals from any real source of input, such as a scanned medium or

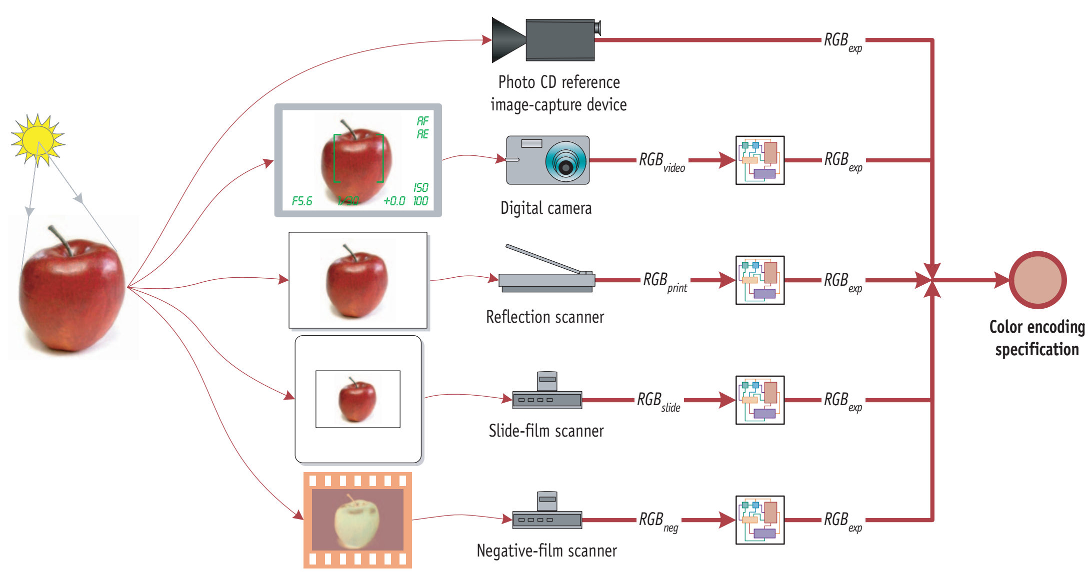


Figure 12.10 Relationship of a reference image-capture device to scene-based color-encoding values. As a result of the input signal processing transformations, encoded values from each input equal those that would have been produced by the reference image-capture device had it captured the same original color stimuli. Note that no transform is required for the reference device itself.

a digital still camera, are processed through input signal processing transforms such that the resulting encoded values equal those that would have been produced by the reference image-capture device had it captured the same original color stimuli.

Practical considerations

If this encoding approach worked perfectly, the encoded values for the stimuli of a given original scene would be identical, regardless of which device or medium actually was used to capture that scene. In *concept*, then, the method should create perfect compatibility among the inputs. In practice, however, there are several complications that must be considered.

The first is that there is a limit as to how closely the values scanned from a given input can be transformed into the values that would have been produced by the reference device. That limit is determined primarily by the spectral sensitivities of the input device or medium used to capture the original scene. The more closely the image-capture spectral sensitivities correspond to a set of color-matching functions, the more closely the encoded values theoretically can match those of the reference device.

Extensive tests showed that colorimetric accuracy was reasonably high for most input media when this theoretical limit was approached, despite the fact that the spectral sensitivities of all tested media differed from color-matching functions. Figure 12.11 shows the colorimetric errors associated with the spectral sensitivities of a representative photographic film. For certain color stimuli, such as those that might be produced by the blue flower and green fabric described in Chapter 1, the colorimetric errors could be somewhat larger. In virtually all cases, however, the errors were considerably smaller than those typically found in the *reproduced* colors of the tested media. Moreover, the test showed that input compatibility among all tested media was excellent when the theoretical limit was approached, even though the spectral sensitivities differed somewhat from medium to medium.

Achieving that degree of compatibility required careful control over the entire process. It also required calibration of the actual medium and equipment that was used and the application of computational techniques that generally would not be available in ordinary practice. Another consideration, then, was whether adequate input compatibility could be achieved under typical operating conditions.

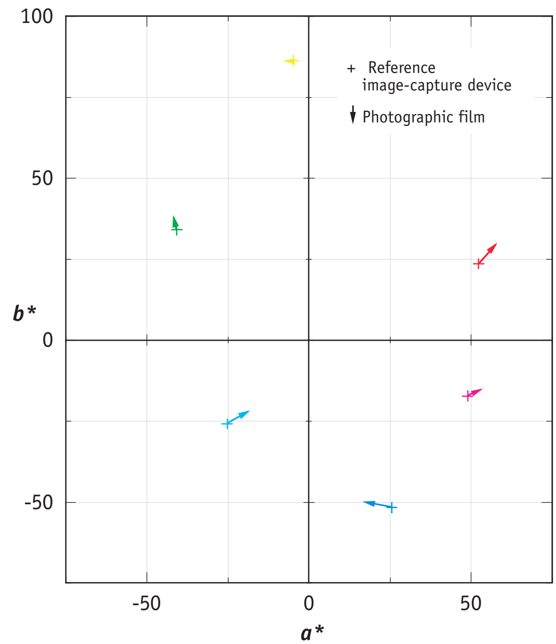


Figure 12.11 Comparison of CIELAB a^*, b^* values determined from reference image-capture device *RGB* values (+ marks) and from a representative photographic film (arrowheads). The differences are due to the departures of the spectral sensitivities of the film from a set of true color-matching functions.

Experiments showed that in order to create a high degree of input compatibility, two things were required. First, custom input signal processing transforms were needed for each input product. In the Photo CD system, these input transforms were called product-specific film terms. Most Photo CD imaging workstations (PIWs), which were used to write Photo CD discs, had more than 100 sets of film terms representing virtually every negative and transparency film available from all major manufacturers. The appropriate film terms were selected, at the start of the scanning process, based on an identification of the specific input film. That identification could be made manually by an operator, or it could be made automatically on workstations equipped with product identification-code readers. However, the use of product-specific film terms alone cannot account for other sources of variability that can occur within a given product. For a photographic film, these might include the following:

- Variations in manufacturing of the film
- Changes in the film after manufacturing, but prior to exposure
- Underexposure or overexposure of the film
- Exposure of the film under nonstandard illumination
- Changes in the unprocessed film and/or latent image after exposure
- Variations and/or deliberate alterations of film chemical processing
- Changes in the film's dye image after processing
- Variations in scanning.

The second thing that was needed, then, was some method to compensate for such variations. For that reason, PIWs were equipped with an automatic lightness adjustment and color-balance adjustment algorithm, called the *Scene Balance Algorithm* (SBA). This algorithm used image-data histograms, pattern recognition, and other analysis techniques to determine sets of red, green, and blue corrections that then were applied as part of the input signal processing.

A scene-based color-encoding method is ideally suited for the application of this algorithm. Because in this method all inputs are encoded on a common basis, it is not necessary to have separate balance algorithms for each different type of input medium. Moreover, the encoding method allows balance corrections to be made in RGB exposure space, which is the space in which most input media variations occur. Underexposure and overexposure errors, for example, are very common, and spectral power distribution differences among scene illuminants also result in relative red, green, and blue exposure variations. A scene-based color-encoding method makes it possible to compensate simply and accurately for these unwanted exposure variations. PIWs also provided a video display of the scanned image. This video preview allowed a scanner operator to make additional adjustments, if necessary, to the scanned image prior to final encoding. A scene-based color-encoding method is of benefit here, again because it allows corrections to be made directly and easily in RGB exposure space. In addition, since the method creates compatibility among the input sources, the same set of software tools can be used to adjust any scanned image, regardless of its origin.

Use of product-specific film terms, together with an SBA and manual adjustments, enhances the consistency of the scanning/encoding process, which in turn results in a very high degree of input compatibility. That raises an interesting philosophical question: how much input compatibility really is *needed*? For that matter, how much compatibility actually is *wanted*?

Degrees of input compatibility

Throughout Part III, the advantages of achieving input compatibility have been emphasized. But *complete* input compatibility may not always be necessary, and it may not always be desirable.

In the case of the Photo CD system, certainly the fundamental incompatibilities of image data from negatives, transparencies, reflection prints, and digital cameras had to be eliminated for the system to work at all. But a scene-based color-encoding method is capable of going farther. At its limit, it can make all encoded images of the same original scene, recorded by any of the input media, virtually *identical.* But should that always be the objective?

By analogy, suppose one were making an audio CD from an old 78 rpm record, and the technical capability existed to process the sound from that record such that the audio CD sounded like a modern digital recording. Should that capability be used, or should the CD reproduce the sound that is characteristic of all 78 rpm records? Stated in more general terms, are there properties of the original recording medium itself that should be retained, or should the effects of all such properties be eliminated whenever possible?

There are no "right" or "wrong" answers to such questions, either for audio applications or for imaging applications. For some advertising and scientific work, it may be desirable to produce image files that are highly accurate colorimetric records of the original scenes, regardless of the actual media used to record those scenes. Similarly, it most often would be preferable for image files produced from transparencies taken of museum paintings to represent the colors of the paintings themselves rather than the colors as reproduced on the transparencies. Applications in which cutting and pasting of images is performed also will work best when images from all input sources are made as much alike as possible.

For other applications, however, an accurate digital record of the photographic reproduction itself might be more desirable. For example, a photographer may have produced an image having very high color saturation by deliberately using a film known to have that type of color reproduction. In that case, it would be preferable to encode the image in a way that retains this high saturation. There is, then, a range of possibilities to consider, from encoding in terms of original scene colors, thereby achieving complete input compatibility, but with no retention of the particular characteristics of the input media, to encoding directly in terms of measured input-media densitometry or colorimetry, thereby achieving complete retention of the particular characteristics of the media, but with no input compatibility.

These two possibilities represent extremes on a continuum of possible trade-offs between input compatibility and retention of specific input-medium characteristics (Figure 12.12). Between these extremes, there are many useful points on the continuum. One of the important advantages of a scene-based color-encoding method is that the point that is reached along the continuum can be readily adjusted.

The degree of input compatibility achieved is controlled by the correspondence between the actual image being scanned and the input signal processing transform being used in the scanner. For example, using transforms specifically developed for every individual input image achieves virtually complete input compatibility. On the other hand, using less specific input signal processing results in somewhat less compatibility, but it retains more of the "personality" characteristics of the particular input medium.

The Photo CD system provided this compatibility control in the form of two different types of input signal processing transforms, which were called *product-specific film terms* and *universal film terms.*

A product-specific input signal processing transform is designed to be used only with the specific input medium for which it was developed. When appropriate product-specific transforms are used for all inputs, as shown in Figure 12.13,

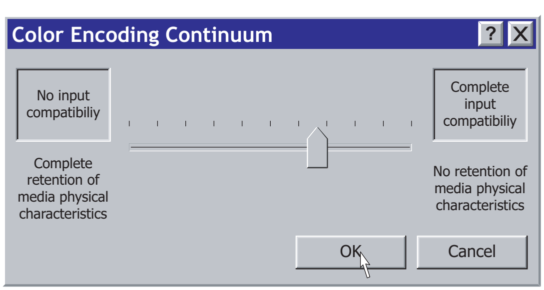


Figure 12.12 There are many useful positions between the two extremes of the color-encoding continuum.

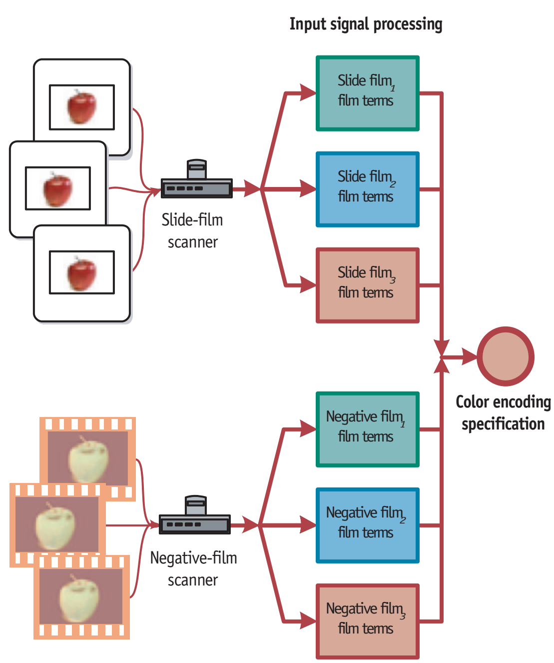


Figure 12.13 Use of product-specific input signal processing transforms in a scene-based color-encoding system.

product-to-product differences among the input media are minimized, and a point quite far to the right on the input-compatibility continuum is reached. Product-specific transforms, rather than universal transforms, should be used when it is important to produce images with a consistent image-to-image look from a mix of different input media. Use of product-specific transforms also results in more seamless compositing of images. Perhaps most importantly, it produces a more accurate extraction of original-scene colorimetry; thus, product-specific transforms should be used when the primary purpose is to encode such colorimetry.

Universal input signal processing transforms are developed for more general use; they can be used for all products of the same basic type (Figure 12.14). For example, one universal transform could be used for all input photographic slide films, another for all input negative films, another for all digital camera JPEG images, etc. A universal transform is based on the characteristics of a reference medium or device of the same basic type as that being input. When universal terms are used, the colorimetric differences of each input from the reference input will be reflected in the color encoding.

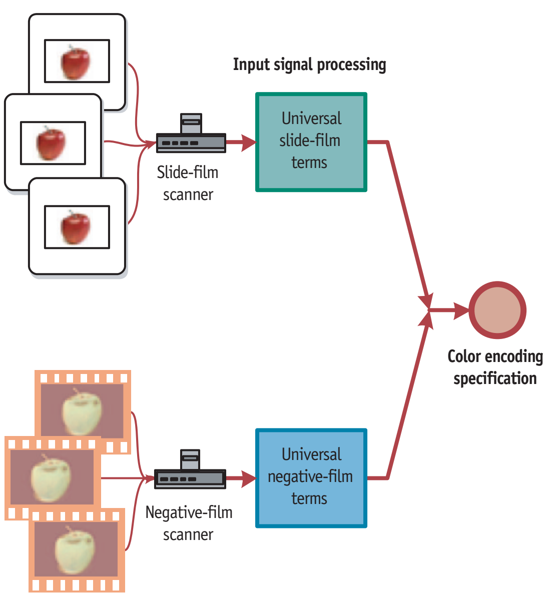


Figure 12.14 Use of universal input signal processing transforms in a scene-based color-encoding system.

For example, a photographic slide film with grayscale contrast higher than that of the reference slide film will produce encoded values representative of that higher contrast. Similarly, slide films with other particular characteristics—such as high or low color saturation, or a tendency to reproduce reds as somewhat orange or somewhat magenta—will produce encoded image files that reflect these characteristics. This option therefore represents a point somewhat more to the left on the input-compatibility continuum, i.e., it favors a greater retention of the characteristics of the particular input medium.

It may be helpful to think of the universal-transform option this way: although it is consistent with the basic definition of encoding in terms of original-scene colors, it results in encoded-scene colors that are altered somewhat according to the particular color-reproduction characteristics of each input film. It is as if the original scenes themselves had been different, having somewhat higher or lower contrast, higher or lower color saturation, more orange or more magenta reds, and so on. When thought about from that perspective, it can be understood why scene-based color-encoded image files created using universal transforms are fundamentally compatible with scene-based color-encoded image files created using product-specific transforms.

A scene-based color-encoding method also allows other encoding possibilities, as shown in Figure 12.15, including the option to encode in terms of *reproduced* color appearance. This option will be discussed in more detail in Part IV, but the basic principle is that reproduced colors can be encoded by transforming them to a "virtual" scene space. This transformation makes these colors compatible with actual scene colors. The advantage of using an original-scene-based approach, even for reproduced colors, is that the method makes it possible to retain *all* of the information that is recordable by any photographic media, all that is recordable by present and anticipated electronic image-capture devices, and all that can be created using computer-generated imaging. This is particularly advantageous in situations

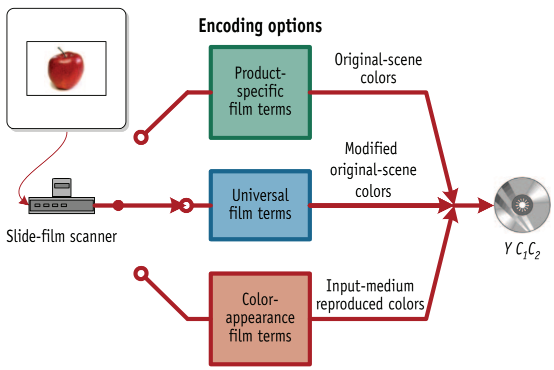


Figure 12.15 A scene-based color-encoding method provides for alternative encoding possibilities. A basic level of input compatibility is achieved in all cases.

where encoded image colors are to be edited and manipulated.

Discussion

The importance of input compatibility has been emphasized throughout this and previous chapters of this part. It was shown that input compatibility is achieved by representing input colors in such a way that encoded values have a single, unambiguous interpretation. In some circumstances, that interpretation can be based directly on standard densitometric or colorimetric measurements. In this chapter, it was shown that the specific needs of the Photo CD system were best met by using a fundamentally different encoding interpretation—one based on the colorimetry of the original scene, as measured by a reference image-capture device.

Also emphasized in this part is the fact that the use of a single interpretation of color values within a system is fundamentally different from simply using a single data metric. Several examples and analogies were presented to demonstrate that the use of a common data metric alone contributes nothing toward solving the basic incompatibility problems caused by disparate input sources.

In fact, to this point no mention at all has been made of *any* particular color-encoding data metric for any of the systems described. This was done deliberately in order to emphasize that the choice of a data metric is unrelated to the achievement of input compatibility. Nevertheless, it often is conceptually and operationally convenient to identify a particular data metric for a given color-encoding method. Determination of the most appropriate data metric for a given method or system is the topic of the next chapter.

Summary of key issues

- Conventional color-encoding methods, such as those based on densitometry or standard CIE colorimetry, cannot meet the objectives of the Photo CD system or of other systems having similar input/output requirements.
- A scene-based color-encoding method can create input compatibility among all types of input media and devices.
- A unique characteristic of scene-based color encoding is that the common color property on which its input compatibility is based can be independent of the color-reproduction characteristics of the input sources.
- When complete input compatibility is desired, scene-based color encoding can represent the colorimetry of the original scene that caused the image to form on each input imaging medium or device.
- The values used to encode the original-scene colorimetry are based on the exposure-factor values that would have been produced by a defined reference image-capture device, had it captured the same original scene.
- Reference image-capture device values are determined using signal processing transformations of image values derived from actual input media and devices.
- Practical color encoding must give consideration to trade-offs between input compatibility and retention of the individual characteristics of input sources.

- Scene-based color encoding allows such considerations to be incorporated as part of the encoding process by the use of product-specific or universal input signal processing transforms.
- When product-specific transforms are used, scene-based color encoding minimizes product-to-product differences among the various inputs.
- When universal transforms are used, scene-based color encoding retains product-to-product differences among inputs of the same basic type.
- Scene-based color encoding also allows encoding in terms of reproduced color appearance. For input compatibility, reproduced colors are transformed to a “virtual” scene space.

13

Color-Encoding Data Metrics

To this point, the first principal attribute of a complete color encoding specification, the color-encoding *method*, has been discussed. The second principal attribute, the color-encoding *data metric*, defines the color space and numerical units in which the encoded color values are expressed. Although it seldom is necessary to restrict a particular encoding method to a single data metric, it often is convenient to do so. Most color-encoding methods do in fact have a particular data metric associated with them.

Many data metrics currently are in use, and attempts frequently are made to identify which of them is "best." Almost every conceivable data metric seems to have its dedicated supporters and its equally dedicated detractors. Our opinion is that the particular requirements of a specific color-imaging system will dictate the most appropriate data metric *for that system*. In practice, systems vary widely in their principal uses, productivity requirements, quality requirements, and computational capabilities. Therefore, there is no single "best" data metric for all systems.

That does not mean, however, that the data metric for a given system can be chosen arbitrarily. Among other things, the data metric affects the efficiency of the color encoding and the requirements for signal processing. The data metric also will affect other factors, such as the visibility of quantization and compression artifacts. In virtually all applications, there are engineering considerations that will favor the use of one data metric over another. In this chapter, the most important of these considerations will be examined.

To do that, the process by which one data metric was developed to meet the specific color-encoding and engineering requirements of a particular system will be described. The data metric is Kodak PhotoYCC color interchange space, which was developed specifically for the Photo CD system. This data metric was chosen for examination here because it is completely defined and because it was developed to meet a particularly challenging set of criteria. Also, since it is a metric developed by the authors, we can offer a first-hand account of the general approach used in the development process. That approach can be used in selecting or developing data metrics that are appropriate for other systems.

System requirements

In order to meet the overall objectives of the Photo CD system, as described in Chapter 12, the data metric for the system had to meet five criteria:

1. The metric had to be capable of encoding a wide gamut of colors and an extensive dynamic range of luminance information.
2. The metric had to provide for image compression incorporating spatial subsampling.

Digital Color Management: Encoding Solutions, Second Edition Edward J. Giorgianni and Thomas E. Madden
 ISBN: 978-0-470-51244-9

3. The metric had to encode images such that digital quantization effects would be visually undetectable.
4. The metric had to allow the system to produce photographic-quality hardcopy output images.
5. The metric had to allow the system to produce excellent-quality video images on television monitors, home computers, and computer workstations, using minimal signal processing.

Each of these criteria significantly influenced the final specifications for the PhotoYCC space data metric. The influence that each criterion had will now be described, starting with the video requirements.

Video requirements

For Photo CD discs to produce high-quality images on existing video systems, and to do so without requiring any special monitor adjustments, the output video signals produced by devices using those discs had to conform closely to industry standards for existing devices that produce output video signals.

It might seem logical, then, to encode image data on Photo CD discs directly according to those standards. That would minimize the amount of required output signal processing, which should translate to the simplest, fastest, and least expensive generation of output video signals. However, there is a fundamental problem with that approach: it is inconsistent with the color-gamut requirements for the system.

Color-gamut requirements

Figure 13.1a shows the chromaticity locations of a typical set of video primaries, as defined by the chromaticities of the light emitted by the red, green, and blue phosphors of a representative video display. To a first approximation, the gamut of colors that can be displayed using these primaries is indicated by the triangle formed by connecting the chromaticity coordinates of the three primaries. (This is only an approximation because the actual color gamut is three dimensional, and the chromaticity limits of the gamut will differ at different luminance-factor levels.) The triangle also indicates the chromaticity

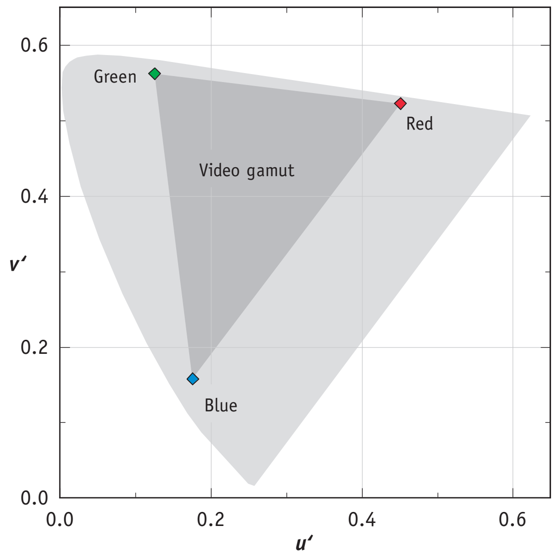


Figure 13.1a CIE u', v' chromaticity diagram showing the chromaticity locations of a representative set of red, green, and blue video primaries.

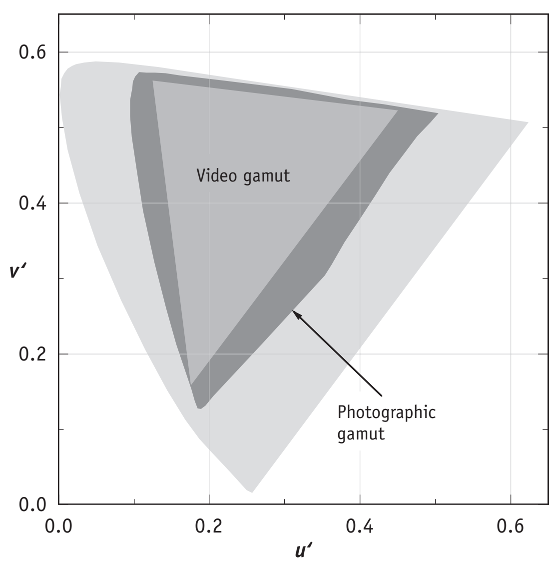


Figure 13.1b Chromaticity boundaries for colors that can be produced using a variety of different photographic CMY image-forming dyes, compared to those for an encoding data metric strictly based on existing video standards.

Figure 13.2 The right half of this represents the color of an original photograph. The left half represents a restriction to a video color gamut. Note the resulting loss of color saturation in the cyan.

boundaries of the color gamut that would result if an encoding data metric based strictly on existing video standards were used.

The chromaticity boundaries for colors that can be produced using a variety of different photographic CMY image-forming dyes, on both reflection and transmission supports, are shown in Figure 13.1b. Note that many of the photographic colors are outside the displayable color gamut of the video system. Consequently, hardcopy output images of full photographic quality cannot be produced from digital images encoded in terms of a data metric that is restricted to a video display color gamut. This is illustrated in Figure 13.2, where the right half of the image represents the color of the original photograph and the left half represents a restriction to a video display color gamut.

Figure 13.3 shows chromaticity boundaries for a collection of colors we have studied that represent real-world surface colors. This collection includes natural colors as well as manufactured colors produced by various pigments, dyes, and printing inks.

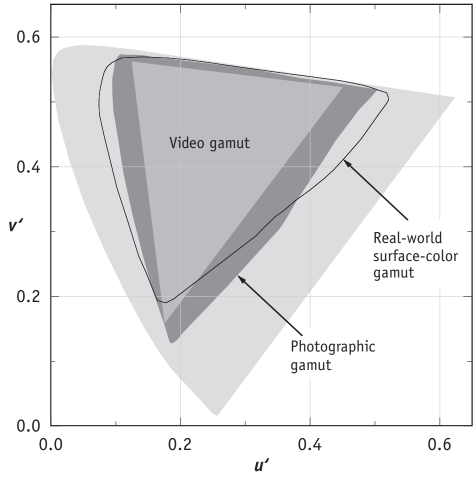


Figure 13.3 Chromaticity boundaries for a collection of colors representing real-world surface colors, compared to those for photographic dyes and those for an encoding data metric based strictly on existing video standards.

Because many of these colors lie well outside the video display gamut, they also cannot be encoded accurately in—or be reproduced from—a data metric restricted to a video display color gamut.

So there was a conflict. If Photo CD images were color encoded using a data metric defined strictly according to existing video standards, the output signal processing required to produce video signals would be simple and fast. But, as has just been shown, the color gamut of that metric would not have included all photographic and real-world colors.

At this point, it seemed that the only option was to use a "device-independent" data metric based on *imaginary* primaries, such as CIE *XYZ* primaries. There would be a significant price to pay in signal processing speed and complexity, but because these primaries easily encompass all real-world colors (Figure 13.4), at least the color-gamut restrictions of a metric based on real device primaries would be avoided. But there was a much better alternative.

The apparent conflict could be resolved by staying with a "device-dependent" metric based on video primaries *if the boundaries of that metric could be extended*. We did this by basing the data metric on the *RGB* tristimulus values of the Photo CD reference image-capture device, which was described in Chapter 12, and by specifying the capture properties for this device such that it became a theoretically perfect video camera.

One "theoretically perfect" characteristic of the reference device is its set of red, green, and blue spectral responsivities (Figure 13.5). We defined these responsivities to be equivalent to the color-matching functions for a set of red, green, and blue video

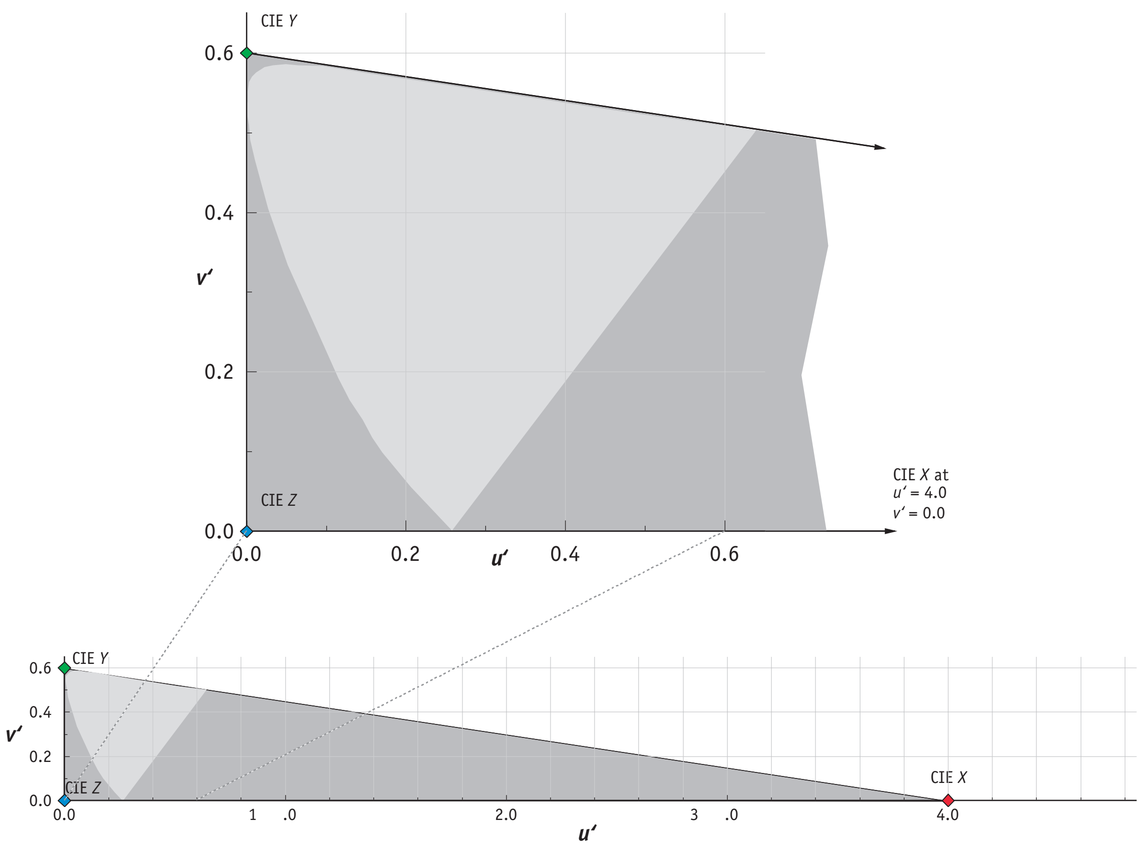


Figure 13.4 Chromaticity boundaries of CIE *XYZ* primaries.

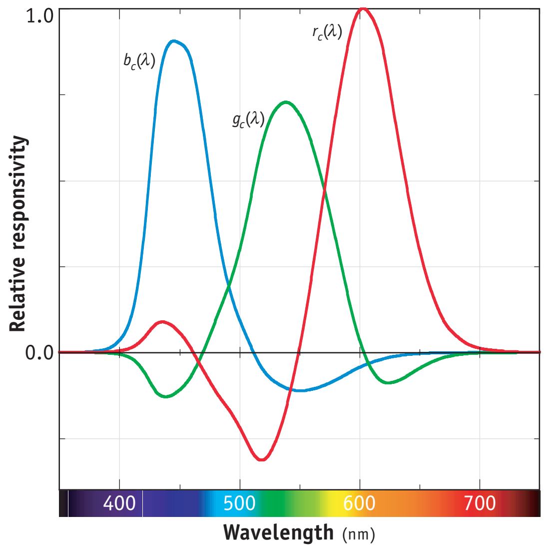


Figure 13.5 The red, green, and blue spectral responsivities of the Photo CD reference image-capture device.

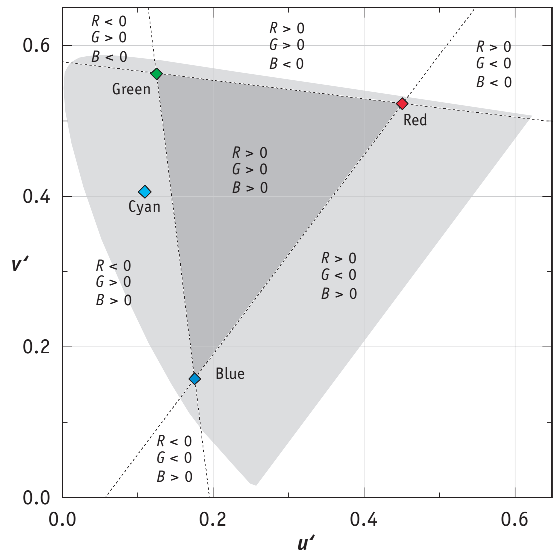


Figure 13.6 Colors outside the displayable color gamut of a video system can be represented by combinations of positive and negative *RGB* values. The indicated cyan color can be represented by values that are positive for green and for blue, but negative for red.

primaries. In particular, we specified the video primaries to be those of CCIR Recommendation 709, *The HDTV Standard for the Studio and for International Programme Exchange.* This recommendation is now ITU-R BT.709, which often is referred to informally as Rec. 709.

We chose to use these primaries because they were consistent with phosphors typically used in the CRTs of modern televisions and computer monitors. Note that since the red, green, and blue video primaries are physically realizable, their corresponding spectral responsivities are not—they are negative at certain wavelengths. But that was not a problem. There was no need to actually *build* this device; it was needed only as a mathematical reference for the data metric.

The advantage of using these particular responsivities is that *positive RGB* exposure-factor values of the reference image-capture device are equivalent to signal values that would be produced by an *actual* video camera that conformed to Rec. 709 specifications. So very minimal signal processing is required to produce standard video signals from data encoded in terms of reference device *RGB* values. Unlike an actual video camera, however, the hypothetical reference image-capture device also is capable of forming *negative* exposure-factor values. These values make it possible to represent colors *outside* the displayable color gamut of a video system (Figure 13.6).

For example, the cyan color indicated by the ◇ mark in Figure 13.6 would be represented by exposure-factor values that are positive for both green and blue, but negative for red. The figure also indicates other areas where the chromaticities are such that one or more of the exposure-factor values would be negative.

To summarize the discussion so far, because its color representation is based on a reference device whose spectral responsivities are derived from standard video primaries, the PhotoYCC space data metric provides encoded values that can be transformed easily to values appropriate for video applications. Positive red, green, and blue exposure-factor values of the reference image-capture device are equivalent to signal values that would be produced by a real video camera that conformed to Rec. 709 specifications. But at the same time, because both positive and negative red, green, and blue exposure-factor values are allowed for the theoretical reference device, the chromaticity boundaries of the encodable color gamut are not restricted to those defined by the Rec. 709 video primaries.

Luminance dynamic range requirements

Another consideration for a data metric is the dynamic range of the luminance values that can be represented. Most original scenes contain information covering a broad range of luminances. In many cases, a considerable amount of that information occurs at very high luminances, including luminances above that of a perfect white diffuse reflector in the principal subject area of a scene. As discussed in Chapter 7, sources of such information include specular highlights, certain types of diffuse highlights, scene areas that are more highly illuminated than the principal subject area, and fluorescent colors.

Because of the visual importance of this above-white information, most photographic materials are designed to record an extensive dynamic range of luminance information. For example, when normally exposed, photographic transparency films have the capability to record (and discriminate) luminance-factor values up to two times that produced by a perfect white in the principal subject area of an original scene (Figure 13.7). Photographic negatives can record an even greater range of luminance-factor values.

In order to produce photographic-quality output images from scanned photographic input media, then, it is necessary to encode values corresponding to an extensive dynamic range of original-scene luminance information. Note in Figure 13.8a, for example, the loss of highlight information in the center image, produced from encoded values limited to the luminance factor of a scene diffuse white, compared to the top image, produced from the greater dynamic range of luminance-factor values recorded by the original photographic image. The bottom image indicates areas where differences exist between the two images.

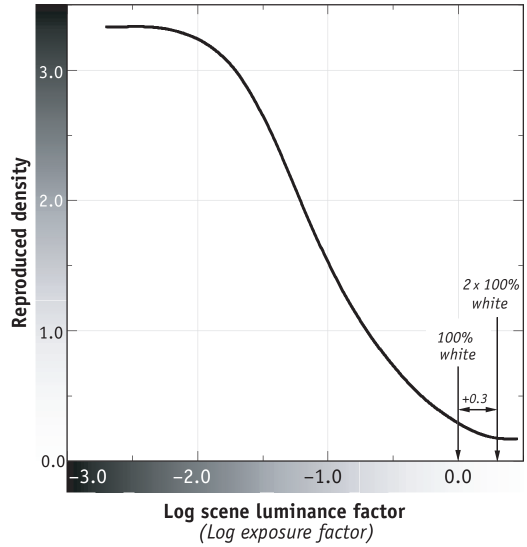


Figure 13.7 Photographic transparency films can record luminance-factor values up to two times greater (+0.30 log) than that produced by a perfect white.

The histogram of Figure 13.8b shows the dynamic range of luminance-factor values of the original scene that was photographed for Figure 13.8a. The histogram relates the number of occurrences (number of pixels) versus luminance-factor value of the original scene. The shaded area indicates scene information corresponding to luminance-factor values above those of perfect diffuse whites within the principal subject area.

Since the reference image-capture device is hypothetical, we could continue to endow it with theoretical capabilities as needed. In this instance, we specified that the dynamic range of its *RGB* exposure-factor values, and therefore its luminance-factor dynamic range, would be unrestricted. When this capability is combined with the capability of forming both positive and negative exposure-factor values, the three-dimensional color gamut of the reference device—and therefore of the *color space* based on that device—also becomes unrestricted.

Figure 13.8a The top image was produced using the full dynamic range of luminance-factor values recorded in the original photograph. The center image was produced from encoded values limited to the luminance factor of a scene diffuse white. Note the loss of highlight information, particularly in the clouds. The bottom image indicates areas where differences exist between the top and center images.

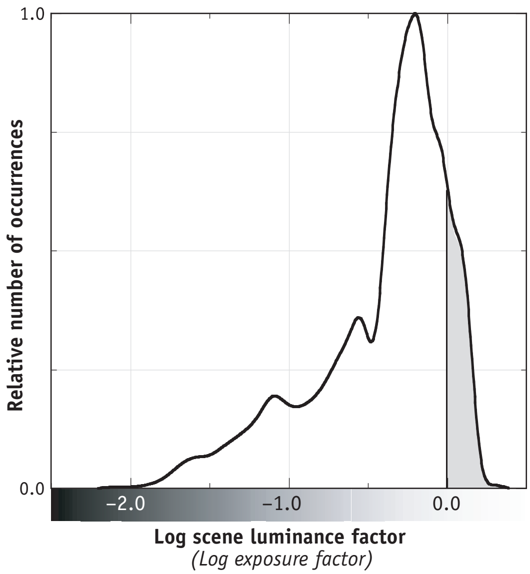


Figure 13.8b Histogram showing the number of occurrences versus luminance-factor value of the original scene shown in Figure 13.8a. The shaded region indicates scene information corresponding to luminance-factor values above those of perfect diffuse whites within the principal subject area.

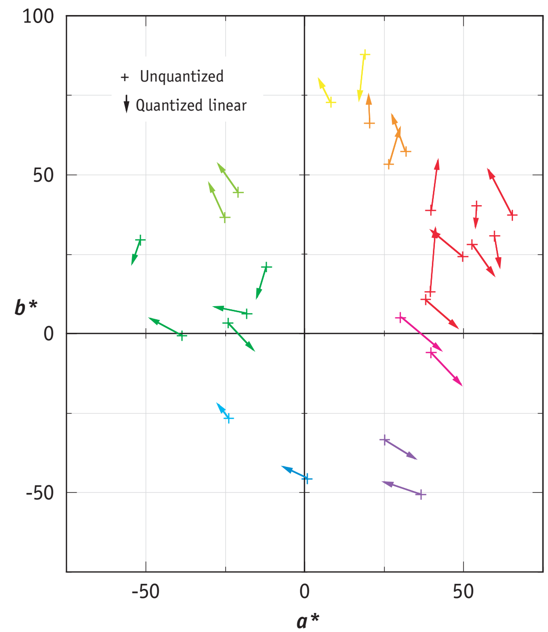


Figure 13.9 Color errors resulting from quantization of linear exposure signals (digitized to 5 bits per color channel). The + marks represent colors produced from unquantized signals; the arrowheads represent colors produced from quantized linear signals.

Digital quantization considerations

Because many Photo CD system applications would use 24-bit color (8 bits per color channel), the potential visibility of digital quantization effects was a concern. If linear *RGB* exposure signals had been digitized, unacceptably visible digital quantization artifacts would have resulted. This is illustrated in Figure 13.9, where the CIELAB a^*, b^* values for a set of original colors and the closest reproductions that can be made from digitized linear exposure signals representing the same original colors are compared. The effect of digitizing linear exposure signals also is demonstrated in the left image of Figure 13.10. *Note that, for clarity and emphasis, the quantization effects have been greatly increased in Figures 13.9 and 13.10 by digitizing to just 5 bits per color channel, rather than the typical 8 bits per channel.*

To minimize color errors and to minimize the visibility of quantization artifacts, linear *RGB* exposure signals first are transformed to nonlinear exposure signals, $R'G'B'$, before any digitization occurs. The nonlinear transformation characteristic, shown in Figure 13.11, is defined by the following sets of equations. The first set describes a basic power relationship between the linear and nonlinear signals. It is applied to linear values greater than or equal to 0.018.

For R, G, $B \geq 0.018$:

$$\begin{aligned} R' &= 1.099R^{0.45} - 0.099 \\ G' &= 1.099G^{0.45} - 0.099 \\ B' &= 1.099B^{0.45} - 0.099 \end{aligned} \tag{13.1a}$$

The second set of equations describes a short linear portion of the transformation. It is applied to linear values less than 0.018.

Figure 13.10 The image on the left was produced from quantized linear exposure signals (digitized to 5 bits per channel). The image on the right was produced from quantized nonlinear exposure signals (digitized to 5 bits per channel). Note the reduction in quantization artifacts in the critical skin-tone regions.

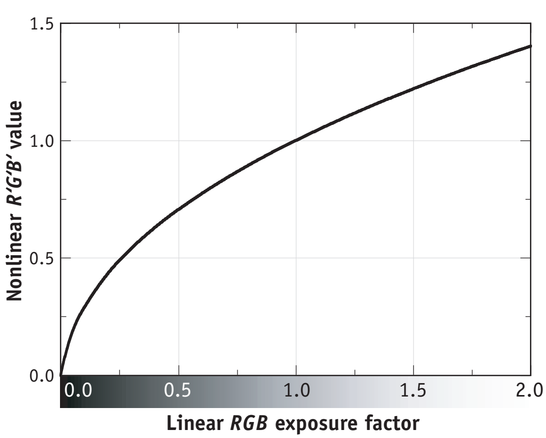


Figure 13.11 Nonlinear transform for converting positive linear exposure signals, *RGB*, to positive nonlinear signals, *R′G′B′*.

For R, G, B < 0.018:

$$\begin{aligned} R' &= 4.5R \\ G' &= 4.5G \\ B' &= 4.5B \end{aligned} \tag{13.1b}$$

This type of transformation ultimately results in a more perceptually uniform distribution of quantization errors, as shown in Figure 13.12 and in the right image of Figure 13.10. In particular, note the significant reduction in quantization artifacts in the skin-tone areas of the photographic image. (*Again, for clarity and emphasis, the quantization shown in these figures is for 5 bits per color channel, rather than the typical 8 bits per channel.*)

Although various other nonlinear transformations could have been used, we chose this particular set of equations for several reasons. First, when applied to positive *RGB* signal values, the transformation corresponds to the video camera *optoelectronic*

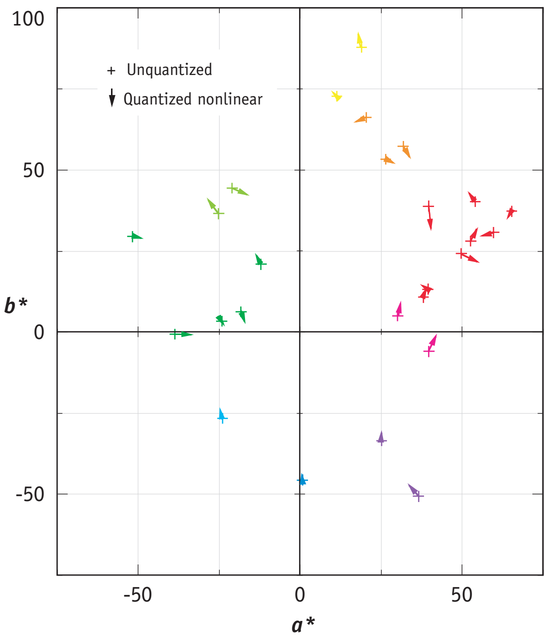


Figure 13.12 Color errors resulting from quantization of nonlinear exposure signals (digitized to 5 bits per channel). Note that the errors are considerably smaller than those shown in Figure 13.9.

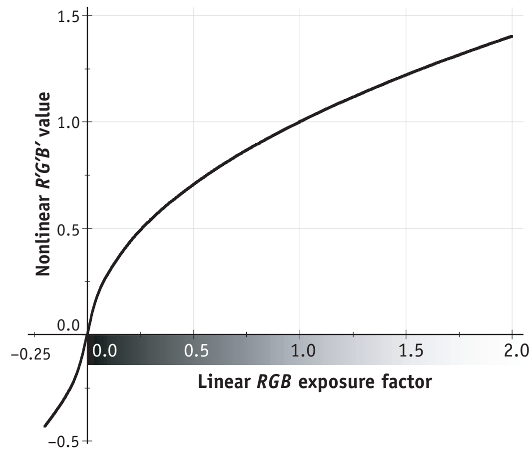


Figure 13.13 Complete nonlinear transform for converting positive and negative linear exposure signals, *RGB*, to positive and negative nonlinear signals, *R′G′B′*.

transfer characteristic specified in Rec. 709. This characteristic defines the relationship between video camera exposures and camera output signal voltages. The use of this particular nonlinear transformation enhances the correspondence of PhotoYCC space values to existing video standards, which further simplifies the signal processing required to produce video-compatible signals from Photo CD discs.

These equations also produced quantization results that were among the best obtained from any of the nonlinear transformations tested. Moreover, these results were retained when images were edited *after* encoding. By comparison, the quantization produced by other nonlinear equations sometimes was quite good in the initial encoding, but problems occurred when image data subsequently were altered to produce modified output images. For example, when some nonlinear transformations were used, quantization effects became noticeable when lighter images were made from encoded values. With other transformations, the effects became noticeable when darker images were made. Images encoded according to the Rec. 709 equations held up well to these types of post-encoding adjustments. This was an important feature, because such adjustments would be an expected part of desktop publishing and other types of editing applications.

In addition to the published Rec. 709 equations, corresponding equations were needed for negative-valued exposure signals in order to preserve the color-gamut capabilities inherent in the color space. After performing a number of calculations and experiments, we determined that the best nonlinear transform for the negative values was a mirror image of the transform used for the positive numbers (Figure 13.13). The resulting set of equations for transforming the complete range of positive and negative linear exposure signals, RGB, to nonlinear signals, $R'G'B'$, is shown below.

For $R, G, B \geq 0.018$:

$$\begin{aligned} R' &= 1.099R^{0.45} - 0.099 \\ G' &= 1.099G^{0.45} - 0.099 \\ B' &= 1.099B^{0.45} - 0.099 \end{aligned} \tag{13.2a}$$

For $R, G, B \leq -0.018$:

$$\begin{aligned} R' &= -1.099\,|R|^{0.45} + 0.099 \\ G' &= -1.099\,|G|^{0.45} + 0.099 \\ B' &= -1.099\,|B|^{0.45} + 0.099 \end{aligned} \tag{13.2b}$$

For $-0.018 < R, G, B < 0.018$:

$$R' = 4.5R$$
$$G' = 4.5G \qquad (13.2c)$$
$$B' = 4.5B$$

Image compression

Before images are written to standard Photo CD discs, they are spatially decomposed into a sequence of image components in a hierarchy of resolutions, ranging from 192 pixels × 128 lines to 3072 pixels × 2048 lines. Higher-resolution images, such as 6144 pixels × 4096 lines and greater, can be stored on Photo CD discs intended for professional applications. In order to store a practical number of images on a disc (more than 100 images on a standard disc), images are stored in a compressed form.

The compression process includes a technique called *chroma subsampling*. In this technique, red, green, and blue signals are converted to an achromatic channel and two color-difference channels. Achromatic information is then stored at full resolution, while color-difference information is stored at a lower resolution. For example, most Photo CD image components are stored in terms of a full-resolution achromatic channel and two color-difference channels that have been spatially subsampled by a factor of two in both the horizontal and vertical dimensions. Therefore, only one value from each of the color-difference channels is used for every four values from the achromatic channel. Because of the nature of the "signal processing" of the human visual system, images produced from full-spatial-resolution achromatic information together with lower-spatial-resolution color-difference information generally are perceived to have the visual quality of full-spatial-resolution color images.

To provide for this compression, the signal processing to the Photo CD system encoding data metric had to include a transformation of *RGB* signals to achromatic and color-difference signals. In theory, this transformation should be performed from *linear RGB* exposure signals (Figure 13.14a). The result of that transformation would be an achromatic *luminance* channel, i.e., a channel corresponding to CIE Y values, and two color-difference *chrominance* channels, i.e., channels corresponding to CIE chromaticity values. However, we found that performing this transformation on linear exposure signals sometimes produced unacceptable subsampling artifacts.

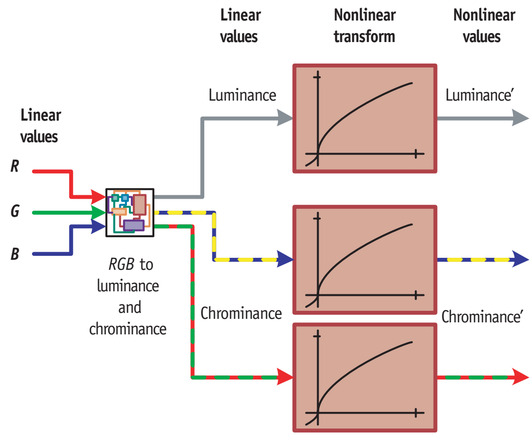


Figure 13.14a This processing arrangement can form true luminance and chrominance channels, but it sometimes generates unacceptable artifacts in subsampled images.

A number of experiments were performed to examine this result, and computational models of the subsampling were developed. This work showed that the visibility of subsampling artifacts would be minimized if the achromatic and color-difference channels were derived from *nonlinear*, rather than linear,

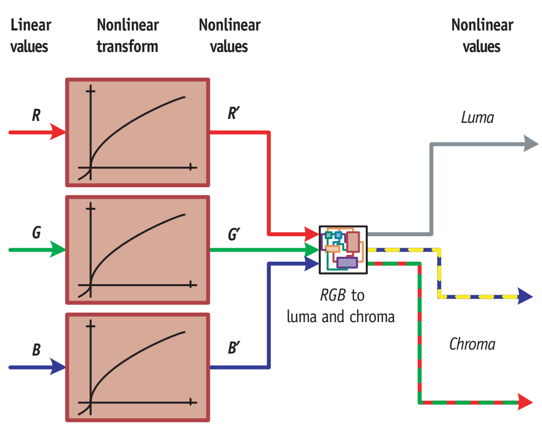


Figure 13.14b This processing arrangement produces images that generally are free of noticeable subsampling artifacts, although it does not form true luminance and chrominance signals. Its outputs are referred to as luma and chroma.

RGB signals (Figure 13.14b). These experiments and computational models also allowed the optimum nonlinear function to be determined.

As it turned out, this optimum nonlinear function was very close to that of Figure 13.13. Since that nonlinearity was needed anyway (to minimize visible quantization effects and to simplify the formation of output video signals), we placed the nonlinear transform *before* the derivation of the achromatic and color-difference signals, as shown in Figure 13.14b. A similar approach has been used for many years in broadcast television, and it continues to be used in HDTV.

One consequence of converting to achromatic and color-difference channels from nonlinear rather than linear signals is that true luminance and chrominance channels are not formed. The resulting channels therefore are referred to as *luma* and *chroma*.

If one were to look at an image formed from either of the two chroma signals, with the luma signal held constant, it would be apparent that the luminance of the chroma image is not perfectly constant. Similarly, although a luma signal, by itself, will always generate neutral signals (equal red, green, and blue), the luma levels for colors are affected to some extent by their chroma levels. None of this would matter if the red, green, and blue signals simply were converted to luma/chroma and then back to red, green, and blue. But when the chroma signals are subsampled, the luminance information they carry also is affected. While a similar situation has had quite noticeable consequences in broadcast television, the effect on Photo CD images generally is undetectable. This result is not surprising, because the level of chroma subsampling in the Photo CD system is relatively small compared to that in conventional broadcast television.

In the PhotoYCC space data metric, the conversion of $R'G'B'$ nonlinear values to luma/chroma values is performed using the following equations:

$$\begin{aligned} Luma &= 0.299R' + 0.587G' + 0.114B' \\ Chroma_1 &= -0.299R' - 0.587G' + 0.886B' \\ Chroma_2 &= 0.701R' - 0.587G' - 0.114B' \end{aligned} \tag{13.3}$$

The coefficients of these equations are derived from NTSC color primaries, not from Rec. 709 primaries. This has relatively little effect on final images, however, because the inverse of each equation is included as part of the output signal processing. The NTSC equations were employed because they are incorporated in most video-signal encoder and decoder circuits and in most video-based software. Their use helps to simplify the conversion of PhotoYCC space values to standard output video signals.

Figure 13.15 shows the chromaticity locations for the PhotoYCC space luma and chroma signals. Note that the chromaticity coordinates of the luma signal are those of D_{65}, which is the reference white for the data metric. This is a significant feature of the metric. What it means is that the extremely important neutrals of an image are represented solely by the luma signal, which is not spatially subsampled in the compression process.

The final metric-related decision that we had to make involved the conversion of luma/chroma signals to digital code values. As stated earlier, the color gamut and luminance dynamic range of the PhotoYCC color space itself are unrestricted. But for digital encoding, upper and lower signal limits must be established when forming digitized luma and chroma values. Selecting such limits requires a compromise of dynamic range versus accuracy. For color gamut, it is important to include as much range as possible in each channel. But if the ranges are too great, errors due to digital quantization will become apparent.

After considerable calculation and experimentation, we specified the following equations for computing PhotoYCC space Y, C_1, and C_2 digital code values (24-bit, 8 bits per channel) from luma/chroma values:

$$\begin{aligned} Y &= \frac{255}{1.402} Luma \\ C_1 &= 111.40 Chroma_1 + 156 \\ C_2 &= 135.64 Chroma_2 + 137 \end{aligned} \tag{13.4}$$

The values for the coefficients and constants of the C_1 and C_2 equations are such that virtually all surface colors, including those found in nature and those produced from manufactured colorants, can be encoded. The required color gamut was determined from a number of published studies and from additional research we had performed. The gamut of these real-world colors is not symmetrical with respect to neutrals. That is why the C_1 and C_2 equations are not centered on a value of 128 and why

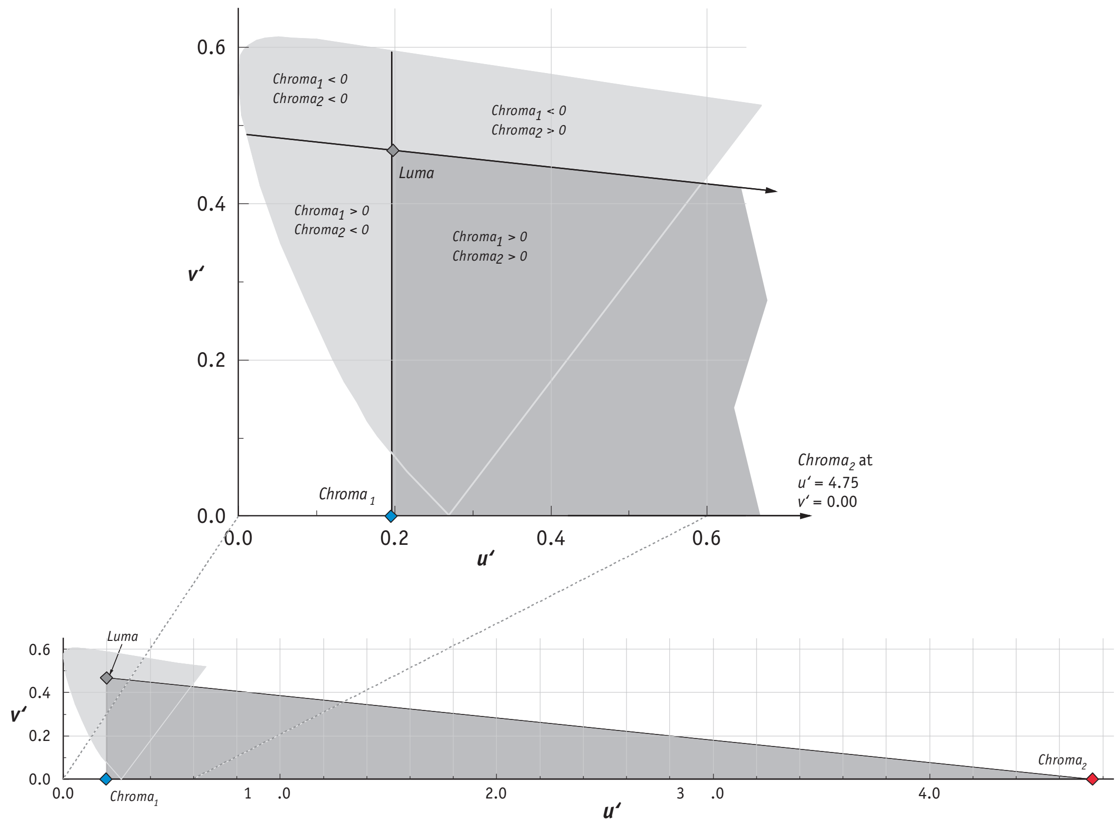


Figure 13.15 Chromaticity locations for the PhotoYCC space luma and chroma signals.

different constants and coefficients are used in those equations.

The practical range of luminance-factor values that can be encoded according to Equations (13.4) extends from 0.0057 to 2.00, i.e., to twice the luminance factor of a perfect white. The total dynamic range therefore corresponds to an approximately 350:1 ratio of luminances, or a log luminance-factor range of about 2.5. This range is sufficient to retain the dynamic range of scene luminances that can be recorded by most electronic cameras and many photographic media.

Implementation

The digital color encoding that has been described here is implemented in the Photo CD system according to the following steps (Figure 13.16):

1. A photographic image is scanned, and *RGB* transmission (or reflection) density values for each pixel are determined.
2. The *RGB* density values are corrected, according to individual scanner calibration transforms, to correspond to those densities that would have been measured by a mathematically defined reference scanner.
3. The corrected density values are processed through the selected input signal processing transform to determine the *RGB* exposure-factor values recorded by the photographic input image. This processing takes place in two steps. The first step removes crosstalk present in the *RGB* density values due to chemical signal processing in the film and to the unwanted absorptions of the film image-forming dyes, as measured by the

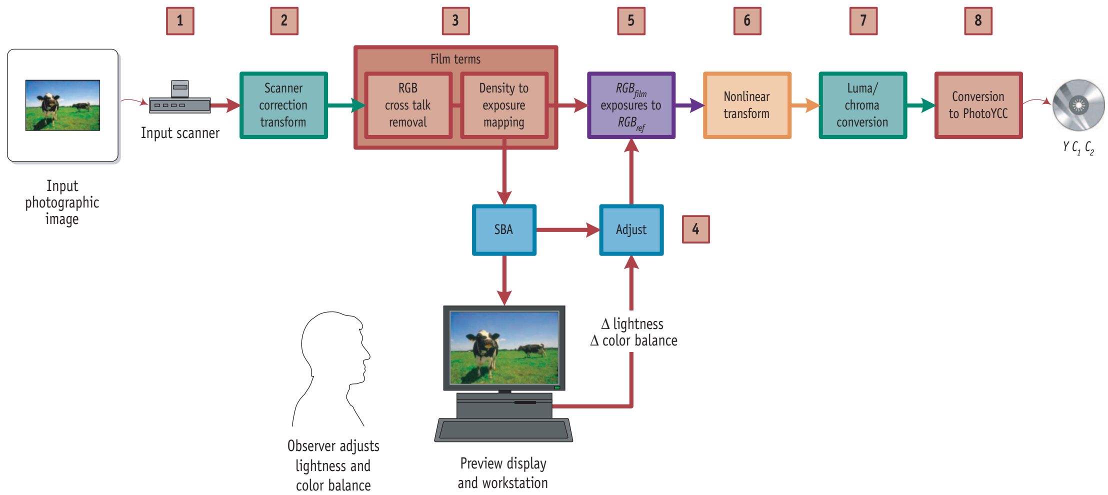


Figure 13.16 Digital color encoding in the Photo CD system.

scanner responsivities. The second step maps the resulting *RGB* density values to photographic *RGB* exposure-factor values.

4. The photographic *RGB* exposure-factor values are corrected for overall lightness and color balance, if desired, by an automatic scene-balance algorithm (SBA) and/or by the scanner operator.

5. The (corrected) photographic *RGB* exposure-factor values next are transformed to equivalent *RGB* tristimulus exposure-factor values for the reference image-capture device.

6. The reference image-capture device *RGB* tristimulus values are transformed to $R'G'B'$ nonlinear values.

7. The $R'G'B'$ values are transformed to luma and chroma values.

8. Finally, the luma and chroma values are transformed to PhotoYCC space Y, C_1, and C_2 digital code values.

Discussion

This description has illustrated how the characteristics of a color-encoding data metric can be tailored to meet the specific requirements of the system on which it will be used. Certainly the Photo CD system data metric would have been very different if, for example, it was meant to minimize the signal processing required to make CMY or CMYK hardcopy images, rather than RGB video images. If image compression had not been required, the metric most likely would have used some form of *RGB*, rather than luma/chroma, color values. If a lesser luminance dynamic range had been required, and/or a greater number of bits had been available, use of a metric that is nonlinear with exposure might not have been necessary.

Although the PhotoYCC space data metric was developed specifically for the Photo CD system, it became widely used for general color interchange for quite some time. That is because its principal features—large color gamut and luminance dynamic range, support for image compression, minimal quantization effects, and fast display to video—are important in a variety of applications. Those features and others will be considered again in later chapters when data metrics are developed for other types of color encoding.

Summary of key issues

- The data metric associated with a color encoding specification defines the color space and the numerical units in which encoded color values are expressed.

- The particular requirements of a specific color-imaging system will dictate the most appropriate data metric for that system.
- Because such requirements and system capabilities vary from system to system, there is no single "best" data metric for all color-imaging systems.
- The PhotoYCC space data metric of the Photo CD system was designed to encode a gamut of colors and a range of luminances sufficient for photographic-quality output, to provide for image compression incorporating spatial subsampling, to minimize the effects of digital quantization, and to produce excellent-quality video images using minimal signal processing.
- Although the PhotoYCC data metric is based on a set of real primaries, its color gamut and luminance dynamic range are restricted only by limitations imposed by the conversions to digital values.
- The general approach used in developing the PhotoYCC space data metric can be used in selecting or developing data metrics that are appropriate for other systems and applications.

14

Output Signal Processing

In the preceding chapters, several different methods for encoding color images in terms of numerical values were described. Once images are encoded according to such methods, they can be stored in digital form, edited, manipulated, and interchanged among systems. But the ultimate use of encoded color images is to produce hardcopy or softcopy output images.

Producing high-quality output images from digitally encoded color values involves consideration of many of the same factors that were discussed in regard to the input encoding process itself. There also are other considerations that are unique to output signal processing and output image formation. A full discussion of all the relevant output-specific considerations is beyond the scope of this book. However, there are certain aspects of the output imaging process that need to be discussed here because they directly relate to color encoding and because they will be referred to in later discussions on complete color-managed systems.

Generating output images

Generating output images requires the use of output signal processing, which converts encoded color values to output-device digital code values. These code values control output-device drive signals, which in turn control the amounts of the color-forming elements that make up the output image. For example, in the case of an LCD display, *RGB* device code values ultimately determine the intensities of displayed red, green, and blue light by controlling voltages applied across the liquid crystal layer in each pixel (Figure 14.1). Similarly, in a thermal printer, *RGB* device code values ultimately determine the amounts of cyan, magenta, and yellow dyes that are transferred to a transmissive or reflective support (Figure 14.2).

Output signal processing can be thought of as a series of individual transforms that perform three basic functions: colorimetric transformation, gamut adjustment, and output code-value determination (Figure 14.3). In some color-managed systems, these individual transforms can be concatenated (combined) into a single transform. However, that is an implementation issue, as will be discussed in Chapter 23. The *effect* of applying a concatenated transform is fundamentally the same as sequentially applying each of the individual transformations discussed here.

Colorimetric transformation

The first function of output signal processing is to transform encoded values such that the colorimetry directly represented or indirectly implied by those values is modified appropriately for the given output. This type of output signal transformation is required, *regardless of the color-encoding method used*, on all systems that include multiple types of outputs.

Digital Color Management: Encoding Solutions, Second Edition Edward J. Giorgianni and Thomas E. Madden
 ISBN: 978-0-470-51244-9

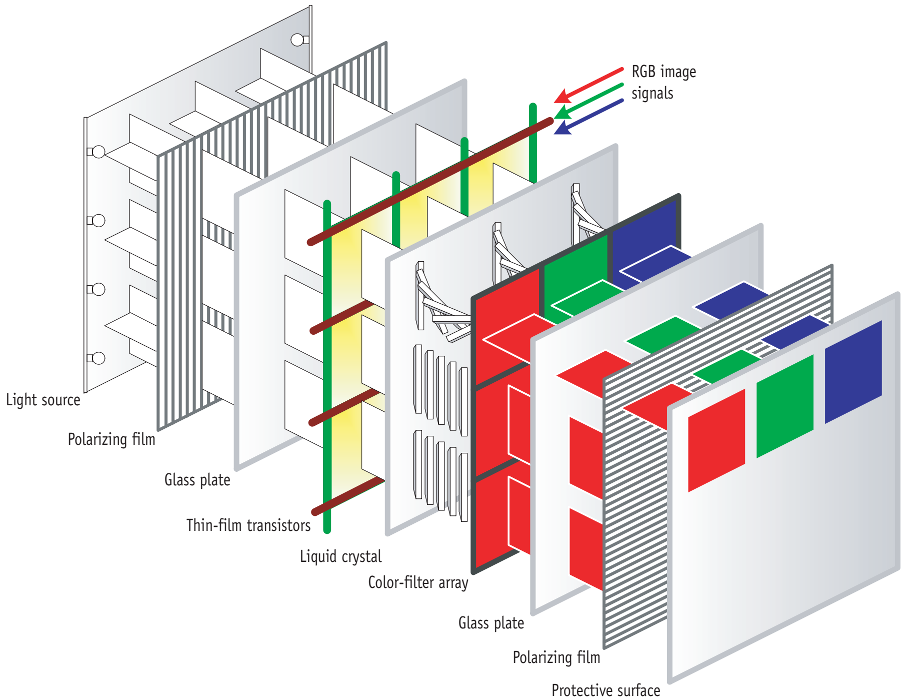


Figure 14.1 Transformation of device RGB code values to red-, green-, and blue-light intensities from an LCD-based display.

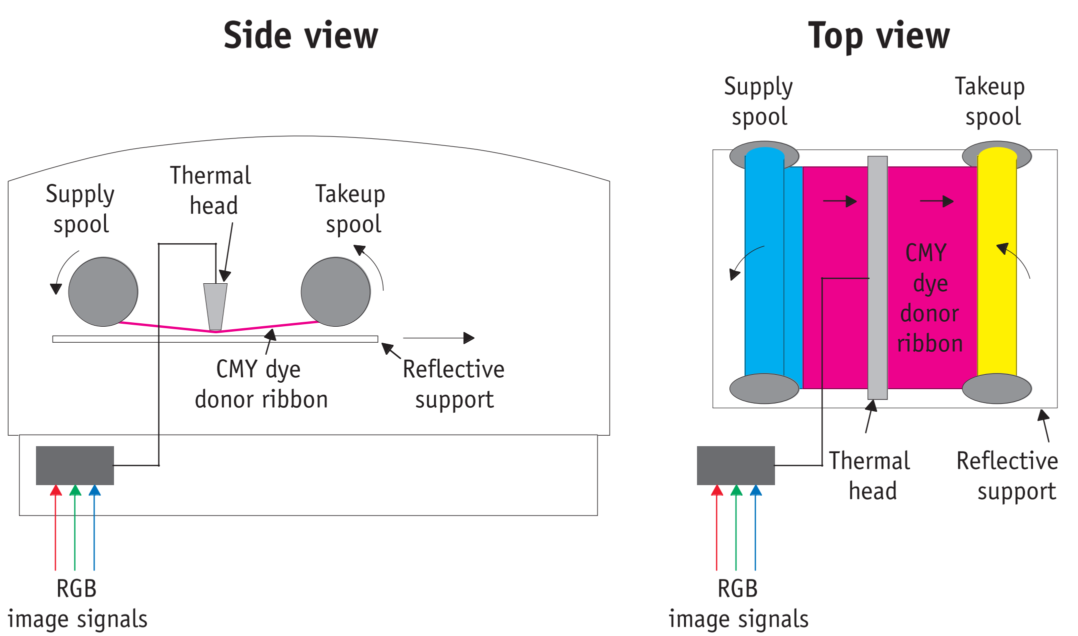


Figure 14.2 Transformation of device RGB code values to thermal print cyan, magenta, and yellow dye amounts.

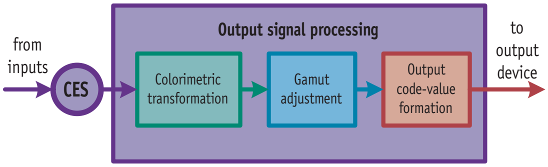


Figure 14.3 The three basic functions of output signal processing.

As discussed in Part II, the colorimetric properties of images on different types of media must be different in order for images to appear appropriately rendered on each medium. Thus, it is impossible to encode in terms of any *single* colorimetric specification, "device independent" or otherwise, that can be used directly for *all* types of output.

The extent and nature of the colorimetric transformation for any given output will depend on the relationship between the method used to create the encoded values and the type of output that will be produced. Thus, for any given encoding method, it is possible that *one* particular type of output will not require a colorimetric transformation. For example, if the encoding method were based on the colorimetry of reflection images viewed under CIE D_{50} illumination, no colorimetric transformation would be required for output to an inkjet printer that produces reflection prints for CIE D_{50} viewing. However, significant transformations of encoded colorimetric values would be required for output to a film writer that produces projection slides or for output to a video monitor.

One way to visualize the need for the output colorimetric transformation is to consider the grayscale characteristics associated with different encoding methods and with different types of outputs. Figure 14.4 shows these characteristics for an original scene (the identity curve (a)), a reflection print (curve (b)), a broadcast video system (curve (c)), and a projection slide film (curve (d)). If the color encoding were expressed in terms of the original scene and the output were to a video display, for example, the output colorimetric transformation would have to include a conversion of curve (a) values to corresponding curve (c) values. If this conversion were not included, the output video images would appear

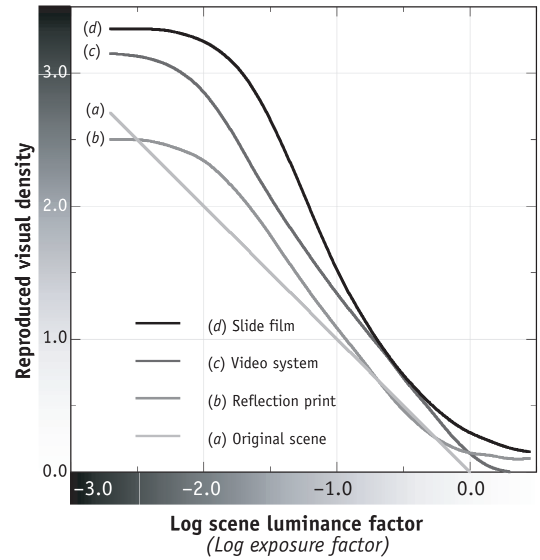


Figure 14.4 Grayscale characteristics for an original scene and three types of output.

too low in luminance contrast. As another example, if the color encoding were expressed in terms of reflection images and the output were to a projection slide film, the output transformation would have to include a conversion of curve (b) values to corresponding curve (d) values. The actual output colorimetric transformations would, of course, involve more than just grayscale values, but the various grayscale relationships help illustrate the basic concept of these transformations.

It also might be helpful to consider output colorimetric transformations in terms of the text-scanning system discussed earlier (Figure 14.5). Color-output devices and media that produce images having different colorimetric properties are analogous to output devices that produce text in different languages. In the case of text, each output requires a *translation* transform from the encoding language to the particular output language. At best, the translation transform for *one* output can be eliminated if the encoding itself is expressed in that output language. In the system shown in the figure, for example, no transform is required for output to Latin. But there is no single encoding language that can go directly, without translation, to *all* of the output languages.

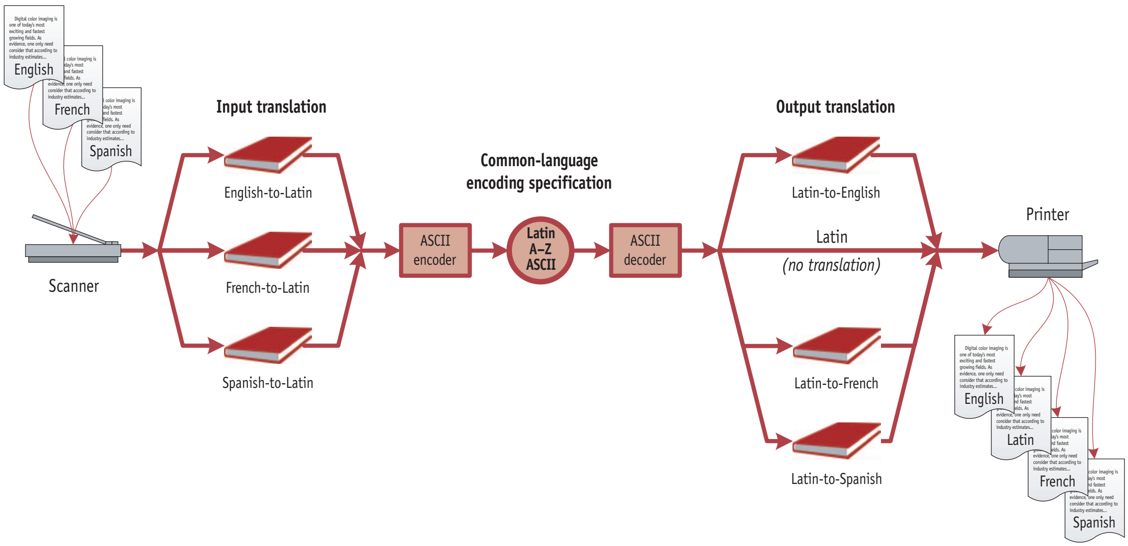


Figure 14.5 Output from a multiple language text-scanning system. There is no single encoding language that can go directly, without translation, to all of the output languages.

In some applications, the colorimetric transformation may include additional modifications of encoded colorimetric values. These modifications are used to deliberately *alter* the color appearance of the output images from that which otherwise would have been produced from the color-encoded values. Such alterations often are used for *simulation*, where one medium is used to emulate the appearance of another. For example, an image might be produced on a computer monitor in a way that simulates what that image would look like if it were produced on a particular type of hardcopy output. Image-editing applications sometimes are used to produce other types of deliberate alterations, such as hue shifts, chroma modifications, and lightness increases or decreases. These changes may be executed by including them in the colorimetric transformation of the output signal processing.

Gamut adjustment

The second function of output signal processing deals with specified output colorimetric values that are not physically reproducible by the actual output medium and/or device. This function, sometimes referred to as *gamut mapping*, replaces out-of-gamut colorimetric values with substitute colorimetric values that can be attained. Some in-gamut values, especially those nearest the gamut boundaries, also may be adjusted so that the substitute values can be accommodated. The criteria used in performing these adjustments and substitutions will vary according to the application for which the output is used.

Some basic concepts of gamut mapping can be illustrated by first examining how such mapping might be applied to monochrome (black and white) images. For example, suppose monochrome images have been encoded in terms of visual density values and are intended for a first medium having a luminance dynamic range of 1000:1. If those values were sent directly to a second output medium having a more limited luminance dynamic range, visual density values beyond that range simply would be clipped (Figure 14.6). The result would be a loss of information in the highlight and shadow regions of the output images, which most likely would be judged to be unacceptable. Another alternative would be to scale the encoded values such that the minimum and maximum visual density values correspond to the

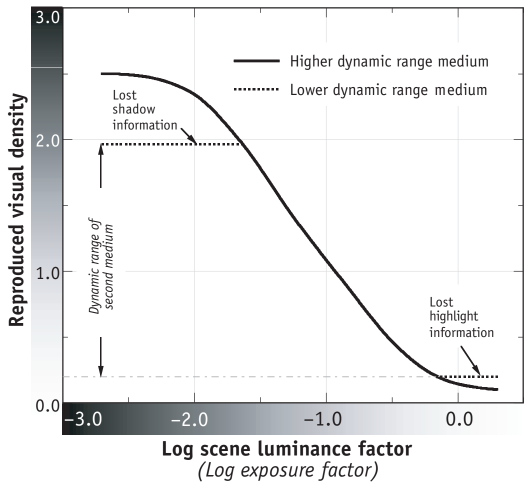


Figure 14.6 Visual density values from a high-luminance, dynamic-range medium will be clipped if sent directly to a second medium having less range.

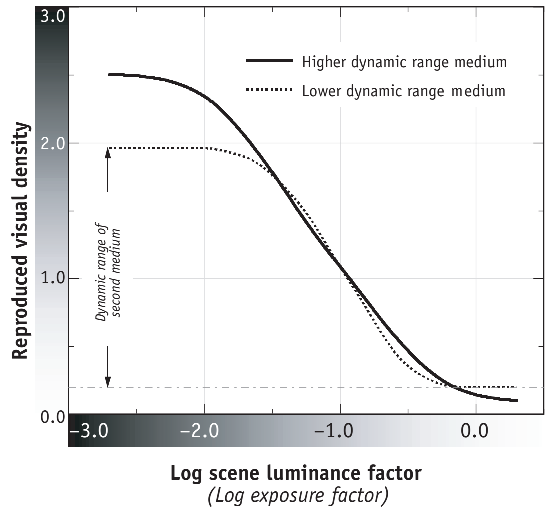


Figure 14.8 Using appropriate signal processing to produce this grayscale on the second medium preserves image contrast and reproduces highlights and shadows with acceptable density values and detail.

respective visual density limits of the second medium (Figure 14.7). However, this process will result in a decrease of overall image contrast, which also is likely to be judged unacceptable.

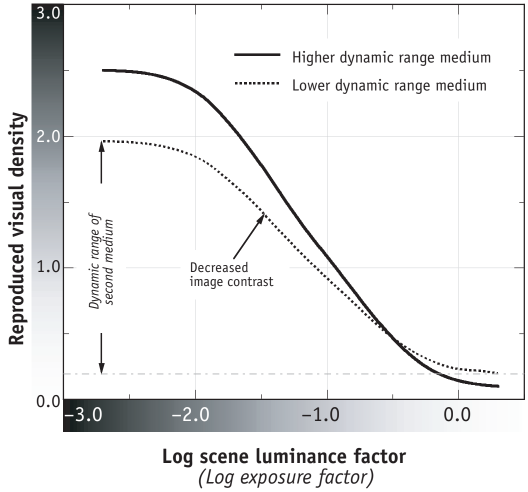


Figure 14.7 Scaling to match the maximum and minimum visual density values of the two example media results in a decrease in overall image contrast in the second medium.

A preferred alternative would be to first determine an optimum system grayscale for the second medium. As discussed in Chapter 6, this grayscale will differ significantly from that of the first medium, even for regions within the visual density range of the second medium, in order to reproduce both highlights and shadows at acceptable density values and with acceptable detail (Figure 14.8). A gamut-mapping transform then can be constructed simply by mapping visual density values for the grayscale of the first medium to those of the second medium at points of corresponding scene luminance-factor value (Figure 14.9). The resulting one-dimensional lookup table (1-D LUT) for this example is shown in Figure 14.10. Note that had the encoding been scene based, the optimum grayscale gamut-mapping LUT for each output medium simply would have corresponded to its respective grayscale characteristic.

Gamut mapping for color images is more complex, of course, and involves more than just visual density values. It is a multidimensional problem, and numerous mapping strategies can be used. For example, the mapping might be based on the minimum distance, in some multidimensional color space, from the out-of-gamut color values to the

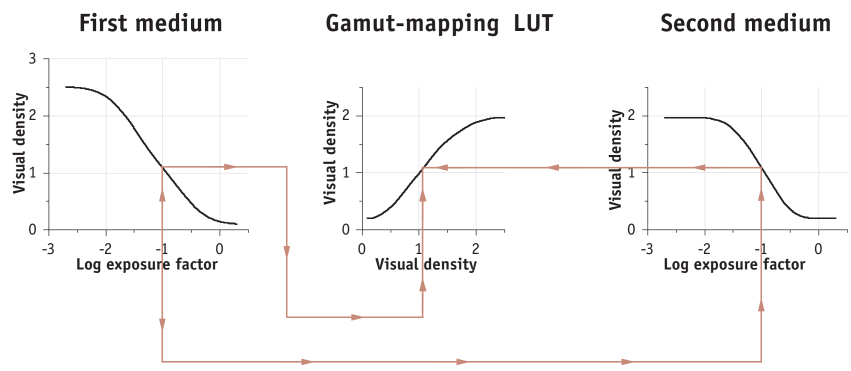


Figure 14.9 A gamut mapping transform can be constructed by mapping visual density values for the grayscale of the first medium to those of the second medium at points of corresponding scene luminance-factor value.

color-gamut boundary of the output medium (Figure 14.11). Alternatively, the mapping might be designed so as to maintain one particular color aspect, such as hue or saturation, of the out-of-gamut colors.

Our experience is that for pictorial images, none of these single-criterion approaches is consistently successful. For example, consider a plane of color space at a particular hue angle corresponding to yellows (Figure 14.12). As the figure shows, mapping the out-of-gamut yellow ♦ according to the minimum distance to the gamut boundary (A) will cause the color to be shifted much darker (lower CIELAB L^*), which most likely will be unacceptable. Mapping the color to maintain chroma (B) or CIELAB C^*/L^*, an approximation of saturation, (C) also will result in the color shifting much darker. Mapping the color to maintain CIELAB L^* (D) will result in the color shifting much lower in chroma. Although (D) likely is more acceptable than any of the first three options, an alternative mapping to a point such as (E), which compromises both lightness and chroma, may be most acceptable in this instance.

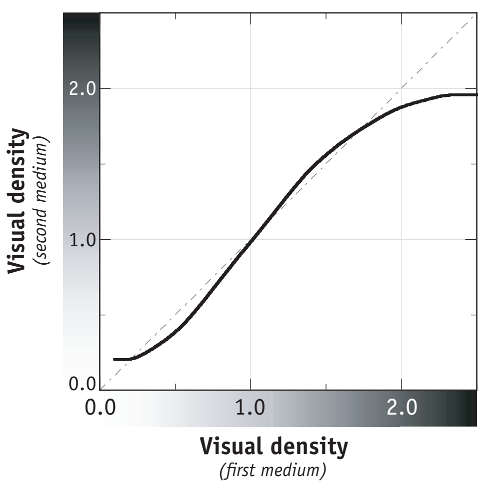


Figure 14.10 A one-dimensional lookup table produced by mapping visual density values for the grayscale of the first example medium to those of the second medium.

The most successful gamut-mapping strategies allow alterations to occur in multiple color aspects and to do so differently in different portions of color space. When applied appropriately, such strategies reduce the visibility of the color mapping and retain as much image detail as possible. This generally requires that some colors within the output medium's gamut be adjusted to "make room" for the remapped colors that had been out of gamut. Doing so retains some discrimination among the remapped colors and preserves at least some of the image detail represented by those colors.

Although such strategies may seem quite complex, acceptable results often can be obtained using relatively straightforward approaches. One such approach, developed a number of years ago

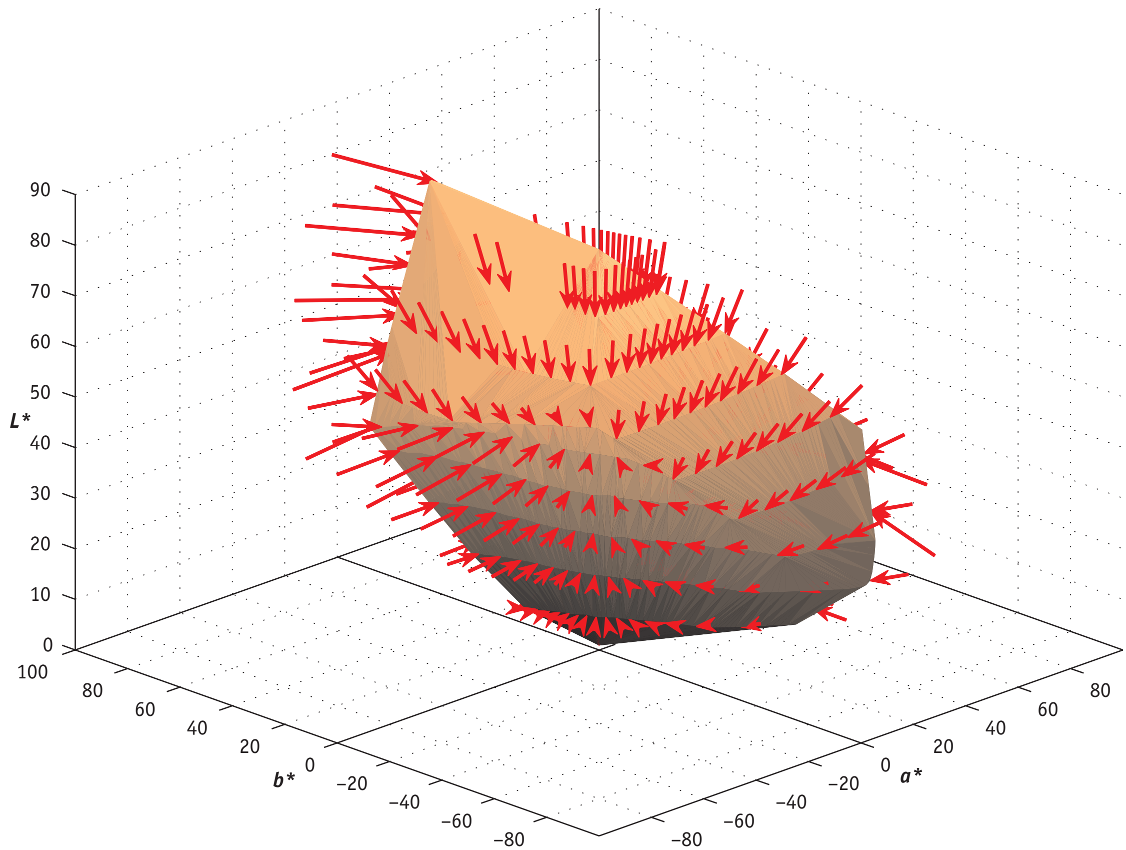


Figure 14.11 The vectors in this diagram illustrate the mapping of color values to a gamut boundary based on the minimum three-dimensional distance.

by the authors, simply extends the one-dimensional mapping previously discussed for monochrome images to three color dimensions (Figure 14.13). The logic of this approach is based on the natural responses of most analog media and devices to input signals.

The inherent responses, i.e., the fundamental relationships of output to input, for most analog devices and media tend to follow some type of S-shaped curve. As a result, a natural form of gamut compression occurs: as signals approach either end of the input range, the rates of change of the corresponding output signals progressively diminish. This can be seen in Figure 14.14, which shows a representative system grayscale characteristic curve and its first derivative. This type of response greatly extends the range over which input signals will produce at least some change in output signals. Note that each of the example 1-D gamut-mapping LUTs that have been shown has had a characteristic that is fundamentally S shaped. This was true whether the mapping was from scene space to an output or from one output to another. As a result, simply applying such LUTs to the individual color signals often can produce color mappings that appear natural and visually pleasing.

The success or failure of this method is highly dependent on the particular color space used for

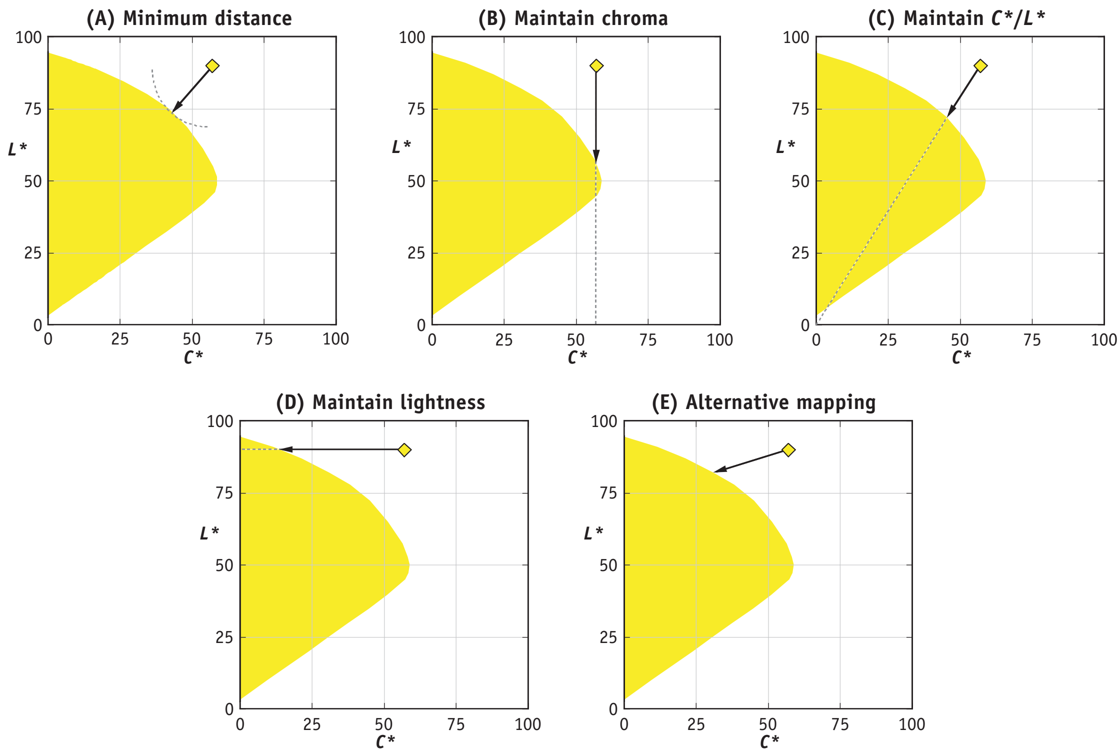


Figure 14.12 Alternative color mappings for an out-of-gamut yellow (♦) according to the minimum distance to the gamut boundary (A), to maintain chroma (B), to maintain chroma/lightness (C), and to maintain lightness (D). Mapping to point (E), which compromises both lightness and chroma, may be more acceptable than any of the single-rule alternatives

its implementation. It is important that the color space be defined such that neutrals are represented by equal values of the three color channels. Doing so prevents the mapping process from introducing hue shifts in neutrals. In addition, the primaries of the color space must be selected appropriately. The choice of primaries will influence how particular colors are affected by the mapping process. In particular, the choice of primaries determines how the hues of individual colors may be affected by the gamut-mapping process. That topic is discussed in detail in the RIMM RGB and ERIMM RGB Color encoding specifications section of Appendix F.

Our experience is that images produced by this straightforward approach generally are more visually pleasing than those produced by many other methods, even those that are much more complex. In some cases, however, where a high degree of gamut mapping is required and/or where the image quality requirements are particularly stringent, the use of three 1-D LUTs alone may not be adequate. In such cases, the basic process can be better implemented using a three-dimensional lookup table (3-D LUT), or a combination of 1-D and 3-D LUTs. Inclusion of a 3-D LUT allows additional corrections to be made for any undesirable hue alterations that might be produced by the 1-D LUTs alone.

Output code-value determination

The third function of output signal processing is to generate output-device code values that will

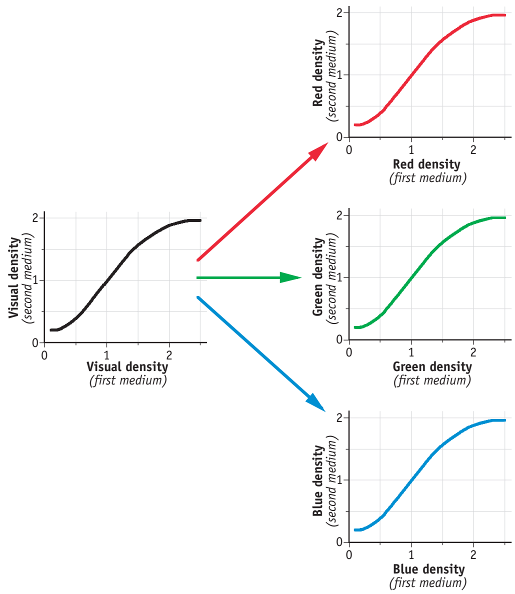


Figure 14.13 One approach to color-gamut mapping is simply to extend the one-dimensional mapping used for monochrome images to three color dimensions.

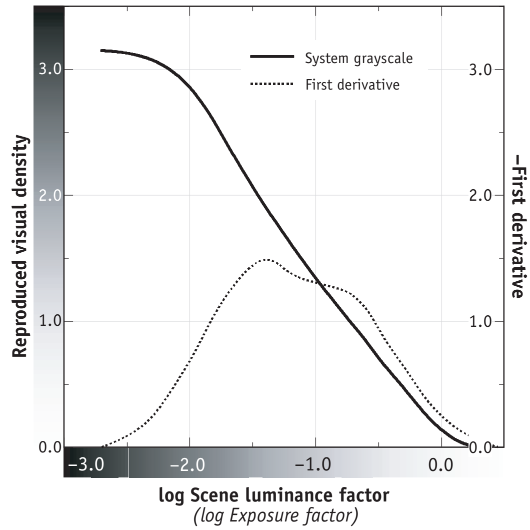


Figure 14.14 A representative system grayscale characteristic and its first derivative. As input luminance-factor values approach either end of the characteristic, the rates of change of the corresponding output visual density values diminish significantly. The range over which input signals will produce at least some change in output signals thus is extended.

produce, on the particular output medium and/or device, output colors having the colorimetric values computed from the color encoding, the colorimetric transformation, and the gamut-adjustment transformation. Derivation of a transformation required for this last function generally consists of two steps: output characterization and output calibration. These terms, which were briefly described earlier, sometimes are used as if they are interchangeable; however, their actual meanings are quite different.

Characterization is a procedure for defining the colorimetric characteristics of a reference device that is representative of the actual devices to be used. A colorimetric output-characterization transform can be developed from data obtained by measuring the colorimetric values of color patches produced on the reference output device from an appropriate array of output-device code values (Figure 14.15). A characterization transform also can be developed by averaging the results of such measurements from a number of representative devices or from a mathematical model of a reference device. In any case, the characterization transform is used to relate desired output colorimetric values to the reference output-device code values required to produce those colorimetric values.

In some circumstances, the use of output characterization transforms alone is feasible. For example, it might be practical to build characterization transforms for each output device in a situation where there are relatively few devices, the devices are relatively stable, and the procedure for building the transforms is fast and economical. In most circumstances, however, it is more practical to build a single reference characterization transform for all devices of the same type and to then provide unique calibration transforms for each individual device (Figure 14.16).

A *calibration* transform corrects for any deviation of a particular device from the reference device on which the characterization transform was

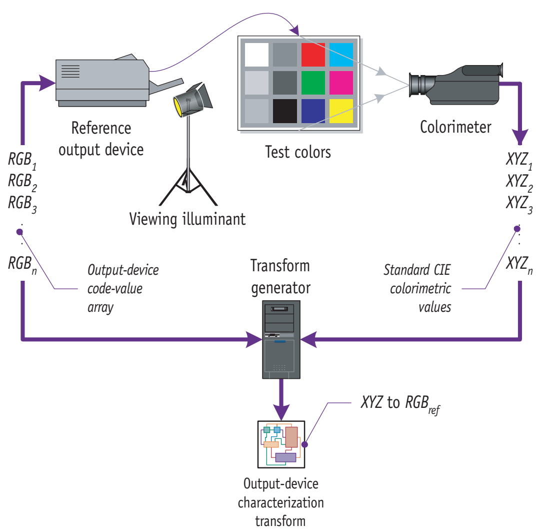


Figure 14.15 Construction of a characterization transform for a hardcopy output device. An array of code values is used to generate a corresponding array of color patches from the output device. The color patches are illuminated, and their colorimetry is measured. A transform relating measured colorimetric values to device code values is then constructed.

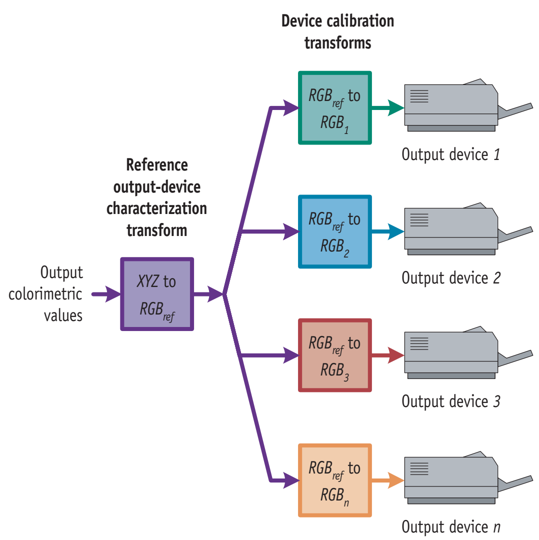


Figure 14.16 Use of a single characterization transform and multiple calibration transforms for multiple outputs.

based. This combined characterization/calibration approach has a number of advantages. In particular, because the bulk of the transformation is performed in the characterization transform, calibration transforms generally can be quite simple. In most cases, they can be derived using a relatively small set of test colors. For many three-color systems, calibration of the grayscale characteristic alone can be sufficient. This generally makes the calibration procedure very fast and inexpensive to perform.

Summary of key issues

- Output signal processing transforms are used to convert color values, encoded in terms of a color encoding specification, to output-device code values. The nature of the conversion depends on the relationship between the encoding method and the colorimetric characteristics of the output medium and/or device.
- Output signal processing transforms generally perform three basic functions: colorimetric transformation, gamut adjustment, and output-device code-value determination.
- A colorimetric transformation is used to convert encoded colorimetric values to modified colorimetric values that are appropriate for the given output. Such transformations are always required in systems supporting multiple types of output media and devices, regardless of the method used for color encoding.
- Gamut-adjustment transforms are used to replace out-of-gamut colorimetric values with substitute values that are physically attainable by the output.
- Output-device code-value determination is achieved by the use of appropriate characterization and calibration transforms.
- An output-characterization transform relates output colorimetric values to output-device code values for a reference device.
- An output-calibration transform compensates for deviations of a particular output device from a reference device.

15

Myths and Misconceptions

The great enemy of the truth is very often not the lie—deliberate, contrived and dishonest—but the myth—persistent, persuasive and unrealistic.

John F. Kennedy

In the preceding chapters, a number of misconceptions regarding color in general and color encoding in particular have been mentioned. These misconceptions are frequently repeated and widely circulated within the industry. Many have become so widespread that they now can be considered "modern myths."

While a few of these myths are relatively harmless, most have had very detrimental effects on the color-imaging industry. In particular, the persistent myths regarding "device-independent" color have seriously interfered with real progress in the areas of color management and color interchange.

It is important that these myths be examined and dispelled before we conclude this part and proceed to Part IV, where the pieces of the digital color-encoding puzzle will be sorted and assembled to form a total picture. For that to be done successfully, there must be no spurious pieces—no leftover myths—still lying around.

In this chapter, then, a number of the more persistent and persuasive-sounding myths related to color encoding will be examined. Some have been mentioned in earlier chapters, others have not. We will attempt to dispel them all by contrasting each with reality (as we see it, of course!).

Myths of device-independent color

Myths usually are created in an attempt to explain what seems otherwise inexplicable, and they often are used to provide support for ideas that people wish to believe. So it is not surprising that numerous myths have arisen regarding the concept of "device-independent" color. After all, color itself certainly seems quite inexplicable at times, and people strongly *want* to believe that color images can be interchanged freely among imaging systems if all images are encoded in a "device-independent" manner.

Unfortunately, despite what its proponents might claim, "device-independent color" most commonly means nothing more than the use of standard colorimetric methods and standard color spaces such as CIE *XYZ*, CIELAB, and CIELUV. As was shown earlier, this form of encoding works only in certain restricted cases, but perhaps because of these occasional successes, several popular myths persist, as follows.

Myth: CIE *XYZ* colorimetric values describe the appearance of color.

Digital Color Management: Encoding Solutions, Second Edition Edward J. Giorgianni and Thomas E. Madden
 ISBN: 978-0-470-51244-9

Reality: This may well be the most common of all color-encoding myths. As discussed earlier, CIE colorimetry was designed for quantifying the trichromatic characteristics of color stimuli. But standard CIE colorimetry alone does *not* specify the color *appearance* of those stimuli. Standard CIE colorimetry emulates the image-capture stage of human vision, but it does not emulate the mental signal processing and image-formation stages that ultimately result in visual perceptions. A specification of color appearance requires the use of advanced forms of colorimetry that account for observer adaptation and other factors. Practical ways of doing that will be discussed in Part IV.

A related prevalent notion is that if CIE *XYZ* values do not describe color appearance, then conversions of those values to other CIE-recommended color spaces such as CIE 1976 $L^* a^* b^*$ (CIELAB) will produce values that do directly relate to color appearance. That would be wonderful if it were true. Unfortunately, it is not. The real purpose of such conversions is to produce colorimetric values such that observed equal color *differences* among pairs of stimuli are represented in a reasonably uniform manner throughout the color space. This is quite useful for specifying colorimetric tolerances because a computed colorimetric difference, such as a given CIELAB ΔE value, will be similarly noticeable regardless of where the involved stimuli are in the color space. But a specification of the location of a stimulus in CIELAB or a similar color space is not a description of its color appearance.

Myth: If CIE colorimetric values alone do not describe the appearance of color, there must be something wrong with the CIE Standard Observer.

Reality: We frequently have encountered situations where color problems were attributed to claimed deficiencies in the CIE Standard Observer specification. In every case we examined, we determined that the problems actually resulted from more mundane causes such as measurement problems, incorrect device calibration, or mathematical errors. In fact, we have *never* encountered a single imaging-related application where using the CIE Standard Observer has been a real problem. This includes extremely demanding applications, such as diagnostic medical imaging, and in situations involving severe metamerism, such as matching colors on hardcopy media with previews generated on CRTs and other types of self-luminous displays. Our experience is that any colorimetric errors introduced by the Standard Observer specification are smaller than normal observer-to-observer differences and far smaller than the colorimetric variability that typically occurs from variations in illuminants, devices, media, calibrations, mathematical transformations, and numerous other system components and operations. To be robust, an imaging system must be designed to deal with these ordinary variations. Therefore any practical system should easily handle any slight errors that might be attributed to the specified responsivities of the CIE Standard Observer.

Myth: Device-independent color encoding allows the input of images from all types of media and devices.

Reality: The *successful* input of images from disparate sources ultimately requires creating input compatibility by encoding in terms of a particular aspect of color that is common among all inputs. Encoding based on standard colorimetric values alone cannot create input compatibility among most types of media and devices.

Myth: Use of a standard, device-independent, CIE colorimetric color space is necessary to provide common image data for interchange among imaging systems.

Reality: Colorimetric *measurements* can provide a basis for color encoding in certain applications. But it is not always desirable—and it is never actually necessary—to express encoded values, derived from those measurements, in terms of standard colorimetric color spaces and numerical units. In many applications, it may be more practical to express colorimetric values in other terms, such as physically realizable device primaries (rather than CIE *XYZ* primaries), color-signal values for a reference output device (a reference HDTV monitor, for example), or tristimulus values for a reference input device (such as a reference digital cinema camera).

Myth: CIE standard colorimetric values are device independent and thus specify actual output colorimetry, independent of which output medium or device is to be used.

Reality: As discussed in Chapter 14, encoded colorimetric values always must be modified appropriately for output to different devices, media, and viewing environments. The use of CIE standard colorimetric values alone does not and cannot eliminate this requirement.

Myth: Only standard CIE representations (color spaces, data metrics, etc.) are device independent.

Reality: This myth stems from a common notion that, for device independence, the characteristics of a color representation must differ from those of *any* device or medium. For example, in one technical presentation we heard, a speaker stated that "PhotoYCC space is not device independent because it is referenced to a device. The fact that the reference device is imaginary is not relevant. It is still a device." That is an interesting criterion for device independence, but using it raises questions about other color representations as well. For example, there are real *XYZ* devices (such as colorimeters). So the speaker's argument would imply that the CIE *XYZ* color space must be even *more* "device dependent" than is PhotoYCC space! We are being facetious, of course, in order to show that it does not make sense to evaluate the device independence of a color representation based solely on its relationship to a device (real or imaginary) or on its conformance to existing standard color spaces. To see why, let us return one last time to our text-scanner analogy.

In that system, Latin—a language that is "independent" of any of the actual languages being used for input or output—was used for encoding. But was it necessary to use an "independent" language? The answer is no. It was necessary to use a *single* language in order to provide input compatibility. But there was no real reason not to encode in one of the actual input or output languages. In fact, there might have been very good reasons to do so.

Suppose, for example, most of the output was to be in English. The system would work most efficiently if the encoding itself also were in English, because no translation would be required for English output. The fact that a "real" language is used is of no consequence, as long as that language has a large enough "gamut" to encode all words of interest. Similarly, there is nothing wrong with representing colors in terms of device-dependent values. Again, there are good reasons to do so, as long as all colors of interest can be expressed in terms of those values.

Based on that reasoning, we suggest this alternative definition for device independence: *A data metric is device independent if the luminance dynamic range and color gamut of its color space are not restricted by an association with a device.*

Although this alternative definition certainly includes color representations used in standard colorimetry, it also includes other representations that may be more practical to implement in actual imaging applications. For example, color values expressed in color spaces based on reference input or output devices often can be used with minimal output signal processing on practical output devices. At the same time, such color spaces can be colorimetrically rigorous, and they can be used to represent an essentially unrestricted color gamut and luminance dynamic range if appropriate encoding techniques are used. For example, the gamut of a color space based on a reference HDTV monitor can be unrestricted if *RGB* intensity values less than zero are allowed. By the definition just given, such color spaces are device independent, yet their values also are "device friendly."

Adaptation myths

Visual adaptation effects are quite complex and difficult to quantify. Perhaps for that reason, a sizable number of myths related to adaptation phenomena have arisen. The following myths directly apply to color encoding.

Myth: Visual adaptation effects are so numerous, complex, and poorly understood that they cannot be accounted for in practical color-encoding methods.

Reality: Sometimes it does seem that there are endless types of visual adaptation phenomena, and it is true that their effects often are quite complex. But from our experience, we believe that there are just three important adaptation phenomena to be concerned with in most imaging applications: chromatic adaptation, general-brightness adaptation, and lateral-brightness adaptation. Although ongoing work continues to increase the understanding of these effects, the present level of knowledge is such that they can be accounted for reasonably well for most imaging purposes. In Part IV, a relatively simple, practical, appearance-based color-encoding method that accounts for these three perceptual effects will be described.

Myth: A chromatic adaptation transform converts the colorimetry measured under one viewing illuminant to the colorimetry that will result (or would have resulted) under a different viewing illuminant.

Reality: This myth is widespread in the industry and so is particularly important to address. A chromatic adaptation transformation, say from D_{50} to D_{65}, *does not* convert D_{50} tristimulus values to the tristimulus values that would have been formed if a D_{65} illuminant had been used instead of a D_{50} illuminant. What the transformation *does* do is determine the corresponding tristimulus values for a color stimulus that would produce, for a standard observer chromatically adapted to the chromaticity of D_{65}, a visual match to an original color stimulus viewed by a standard observer who is chromatically adapted to the chromaticity of D_{50}.

Contrary to a common but incorrect assumption, the full illuminant spectral power distributions of the two adaptation conditions are *not* required—nor are they used—in the chromatic adaptation transformation process. As shown in Appendix D, only the *chromaticities* of the two *observer chromatic adaptive states* are required in the derivation of a chromatic adaptation transform. That fact alone should make it clear that the transformation does not determine colorimetric values that would be produced if the spectral power of the illuminant were changed.

One factor that may contribute to the confusion on this subject is that in some special circumstances it *is* possible to determine, from colorimetric values, the colorimetry that would result under a different illuminant. However, doing so does *not* constitute a chromatic adaptation transformation. In addition, the determination can be performed accurately from trichromatic values only for colors made from three (or fewer) colorants of known spectral characteristics. For example, if the colorimetry (or densitometry) of a color photograph is measured, and if the spectral sensitivities of the measuring device, the spectral power of the measuring illuminant, the spectral absorption characteristics of the CMY dyes, and the spectral absorption and other optical characteristics of the reflection support *all* are known, the spectral reflectance of each measured color can be calculated from the measured colorimetry or densitometry. This is possible because there is one, and only one, spectral reflectance that can be made from those particular CMY dyes, on that particular support, and under that particular illuminant, that will result in a given triad of colorimetric or densitometric values. Once the reflectance is determined, the colorimetry for that reflectance plus *any* illuminant spectral power distribution can be computed using ordinary colorimetric calculations. It is important to emphasize that this technique, which is used in some input scanners, works only because the original colorimetry is derived from a known set of three colorants. It also should be emphasized once again that this type of computation is *not* a chromatic adaptation transformation.

If the distinction between a recomputation of tristimulus values using a different illuminant spectral power distribution versus a transformation of tristimulus values from one state of observer chromatic adaptation to another still is not clear, it might be helpful to consider the case of metameric colors. Assume there are two reflection color patches that are metameric matches under a reference illuminant *A* but not under a spectrally different illuminant *B*. Also assume that the observer is fully adapted to the respective chromaticities of illuminants *A* and *B* when viewing the color patches under the corresponding illuminant. The following statements then would apply:

- The tristimulus values for the two patches would be the same when computed using the spectral power distribution of reference illuminant *A*. This must be true, by definition, since it was stated that the patches are metameric for that illuminant.
- If tristimulus values for both patches were computed using the spectral power distribution of illuminant *B*, they would be different from those computed using the spectral power distribution of illuminant *A*. Moreover, when computed using illuminant *B*, the sets of tristimulus values for the two patches would be different *from each other*. That outcome is consistent with the statement that the two patches are not metameric under illuminant *B*.
- If instead a chromatic adaptation transform were applied to the reference illuminant *A* tristimulus values to produce visually corresponding tristimulus values for an observer chromatically adapted to the chromaticity of illuminant *B*, the transformed values would, of course, be different from the starting values. However, what is important here is that the sets of transformed tristimulus values *would be the same for both color patches*. This must be the case since the same transformation is applied to the illuminant *A* tristimulus values, which were the same for the two patches.

This outcome is consistent with the fundamental intent of a chromatic adaptation transformation process. In this example, the transformation determines the particular colorimetric values necessary for an observer adapted to the chromaticity of illuminant *B* to see the patches as they appeared under the reference illuminant *A*. The two color patches visually matched under the reference illuminant. In order to appropriately represent that fact, the transformed tristimulus values *should* be the same for both patches.

Myth: An observer adapts to the brightest color of an image, and that color is perceived as a perfect white.

Reality: This myth has been the cause of many color interchange problems within the imaging industry. As discussed earlier, the amount of adaptation to an image will depend on the conditions in which that image is viewed. There can be a very significant amount of adaptation to projected and self-luminous images, especially when they are viewed in darkened environments.

However, even in these cases, what is perceived as a perfect white seldom, if ever, corresponds to the *brightest* color of the image. Experiments have shown that observer chromatic adaptation is controlled primarily by the chromaticity of colors judged to be scene neutrals of about 40 % reflectance. A color that would be judged to be a perfect white would have that chromaticity, and it would have a perceived brightness equivalent to that of an ideal reproduction of a perfect white diffuse reflector in the original scene. The grayscale characteristics of all high-quality imaging media and systems are designed such that images contain areas that appear *brighter* than the reproduction of a perfect white. Such areas create the illusion of being “whiter than white.” They are interpreted appropriately as specular highlights, highly illuminated diffuse objects, etc.; they are *not* perceived as white references—either in brightness or in chromaticity—for other colors.

This myth is even farther from reality for reflection images. The assumption underlying the myth is that observers adapt to the “white” defined by the reflection support, since the brightest areas that can be produced on a reflection image consist of the support alone. In reality, however, reflection images generally are judged as objects within a viewing environment. Under most viewing conditions, there is little or no adaptation to a reflection image itself because the viewing environment contains white objects and other visual cues that serve as references for the general-brightness adaptation and chromatic adaptation of the observer.

It might be useful to consider that if observers really did adapt to the brightest image color, media manufacturers would not bother spending as much time and money as they do trying to make reflection-print support materials as bright and as neutral as possible. In reality, customers *easily* can see differences among supports, even when those differences are so small that they are difficult to measure with the most sensitive instruments available. That would not be the case if observers did indeed adapt to the "white" of the support.

This myth will be discussed in greater detail in Chapter 19, where the problems associated with what has become known in the imaging industry as "relative colorimetry" (or somewhat more appropriately as "media-relative colorimetry") are described.

Scene-based color encoding and data metric myths

Scene-based color encoding is fundamentally different from other methods, and it often is misunderstood. As a consequence, many published explanations of that form of encoding have not been entirely accurate. In particular, most descriptions have not distinguished between the color-encoding method and any associated data metrics. This has led to a number of persistent myths that must be cleared up before we proceed to Part IV, where the scene-based encoding method will be one component of a comprehensive color-management system.

Myth: Colors encoded in terms of scene-space data metric values always represent original scene colors.

Reality: Certainly, scene-based color encoding can be used to represent original scene colors. However, additional operations can be used to transform *reproduced* image colors into scene-space data metric values. When this is done correctly, the resulting images are fully compatible with encoded original-scene images. In fact, when appropriate procedures are used, a scene-based data metric can be used to encode colors from *any* source, regardless of whether that source represents actual original scene colors, modified original scene colors, imaginary original scene colors (e.g., from computer-generated images), reproduced colors, or any other types of colors. This will be discussed in greater detail in Part IV. For now, what is most important is to recognize that scene-space encoding refers to colorimetric specifications of color stimuli viewed under conditions that normally would be associated with live scenes. The *origin* of such stimuli is a completely separate issue.

Myth: Use of product-specific transforms results in the encoding of original-scene colors, whereas use of universal transforms results in the encoding of colors as reproduced by the input medium.

Reality: This myth began years ago as people struggled to understand the scene-space encoding of the newly introduced Photo CD system. Unfortunately, the myth spread rapidly and is now widely believed. It has been repeated at numerous technical conferences and has appeared in several textbooks and other publications. But as described earlier in Chapter 12, the use of *either* product-specific *or* universal transforms results in the encoding of scene-space colorimetry.

The real distinction is that using product-specific transforms for each input will encode actual original-scene colorimetry with greater colorimetric accuracy. For a given input, that accuracy is determined by the accuracy of any input calibration/characterization procedures and by the correspondence of the spectral sensitivities of the input device or medium to a set of visual color-matching functions. When a single universal transform is used for a group of related inputs, the accuracy of the original-scene-colorimetry encoding is further influenced by the extent of the signal processing differences that might exist among those inputs.

Myth: The luminance dynamic range of scene-space encoded values is insufficient for many types of output devices and media, which may have dynamic ranges of 3000:1 or greater.

Reality: There are two misconceptions here. First, luminance dynamic range and color gamut actually are unrestricted in scene-space encoding. Limits are defined only when digital code values are assigned, based on the bit depth available in the encoding data metric. To avoid quantization problems in a 24-bit system (8 bits per channel), for example, the luminance dynamic range might be restricted to perhaps 300:1 (about 2.5 in logarithmic units). A greater scene-space luminance dynamic range can, of course, be encoded in applications that support greater bit depths. The second, more basic misconception is that the dynamic range of scene-space encoding can be compared directly to the dynamic ranges of images used for input or images produced for display by an imaging system. It cannot.

Scene-space values correspond most closely to *input* exposure values of an image-capture device or medium, while values for other forms of color encoding typically correspond to measurements of *output* images produced by a medium or system. The ranges of input and corresponding output values are quite different. Figure 15.1 shows, for example, the relationship of scene-space log luminance-factor values to corresponding reproduced values for a representative imaging system. Note that in this example, a 2.5 dynamic range (about 300:1) of encoded scene-space values is sufficient to fully represent the corresponding range of output values of 3.5 (about 3000:1) that

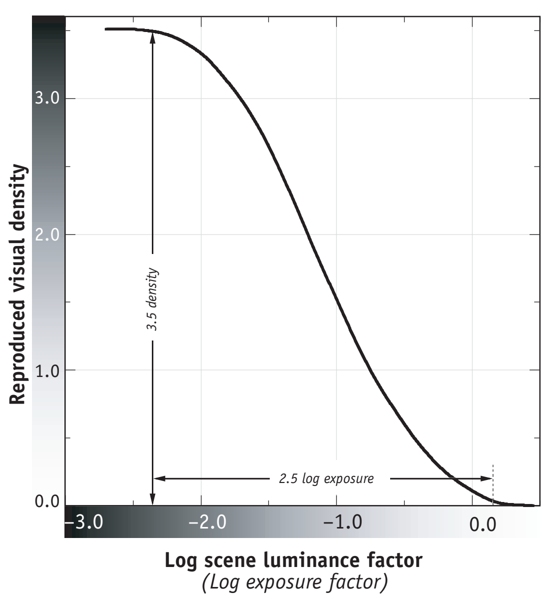


Figure 15.1 Relationship of encoded scene-space values to output values for a representative imaging system.

can be produced by that system. The figure illustrates the appropriate way to compare a scene-space dynamic range to that of a device, medium, or system. A direct comparison of scene-space values to output values is an essentially meaningless "apples and oranges" comparison.

A scanner myth

Densitometric scanners, which were discussed in Chapter 10, and colorimetric scanners, discussed in Chapter 11, share many similarities. However, they have one fundamental distinction: while the physical spectral responsivities of a colorimetric scanner are based on a set of visual color-matching functions, those of a densitometric scanner are not. As a result of that difference, the functionality of the two types of scanners is quite different.

Myth: The functionality of a scanner that incorporates transforms to convert scanned *RGB* values to CIE colorimetric values is identical to that of a colorimetric scanner.

Reality: An RGB scanner will produce accurate CIE colorimetric values, based on a particular light source, *only* when used with a transform that is specifically designed for the particular medium being scanned. If a medium having a different set of colorants is scanned, the transformation likely will be inaccurate. A true colorimetric scanner can measure CIE colorimetric values, again based on a particular light source, for *any* medium. This will be true even for complex media, such as paintings, in which a large number of colorants have been used.

Color-reproduction myths

Myth: The ideal grayscale reproduction should be one-to-one with the original scene.

Myth: The "S-shaped" grayscale characteristic of conventional photographic systems is unfortunate and results from chemical limitations. The grayscale characteristic should be a straight line.

Myth: Ideal color reproduction would exactly duplicate the colorimetry of the original scene.

Reality: These myths share a common theme: the relationship of reproduced colorimetry to original-scene colorimetry. Each myth ignores physical and perceptual factors that must be taken into account in order to make reproductions that best approximate the appearances of original scenes.

When appropriate compensations for viewing flare, general-brightness adaptation, lateral-brightness adaptation, differences in scene and reproduction luminance levels, differences in scene and reproduction dynamic ranges, differences in scene and reproduction color gamuts, color-memory effects, and color preferences all are factored in, the resulting optimum color-reproduction characteristics are quite similar to those of commercial photographic products. Optimum grayscale characteristics are highly nonlinear; they certainly do not have a simple one-to-one relationship to original scene luminance factors. Similarly, optimum color reproduction is not a simple one-to-one colorimetric match to the original scene.

One possible source of confusion on this issue is that original "scenes" can be live scenes, or they themselves can be reproductions. When an imaging system is used to produce a copy of an input that *itself* is a reproduction, that system should produce a one-to-one replication of the colorimetry of that input reproduction, as shown in Figure 15.2. This assumes, of course, that the input reproduction and the output copy are to be viewed under identical conditions. In such cases, the departures from colorimetry required for realistic depictions of live scenes already have been incorporated in the input reproduction. Those departures need not, and should not, be made again in the replication process.

Myth: There are no whites "whiter" than a perfect white diffuse reflector.

Reality: This color-reproduction myth misses the distinction between objects and color stimuli. Most original scenes contain information at luminance levels well above that of a perfect 100 % white diffuse reflecting object in the principal subject area of the scene. This information comes from stimuli produced by specular reflections, by certain types of diffuse reflections, by areas of a scene that are more highly illuminated than the principal subject area, and by certain types of fluorescent colors.

Because of the visual importance of this above-white information, high-quality imaging media and devices are designed to record and display that information. Therefore, digital color-encoding methods and data metrics used with such media and devices

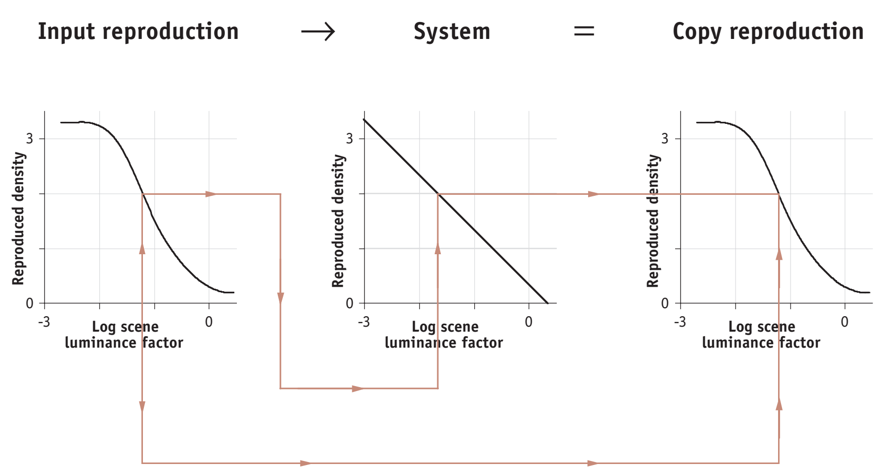


Figure 15.2 In the special case where an imaging system is used to produce a copy of an input image that itself is a reproduction, the system should produce a one-to-one replication of the colorimetry of the input image.

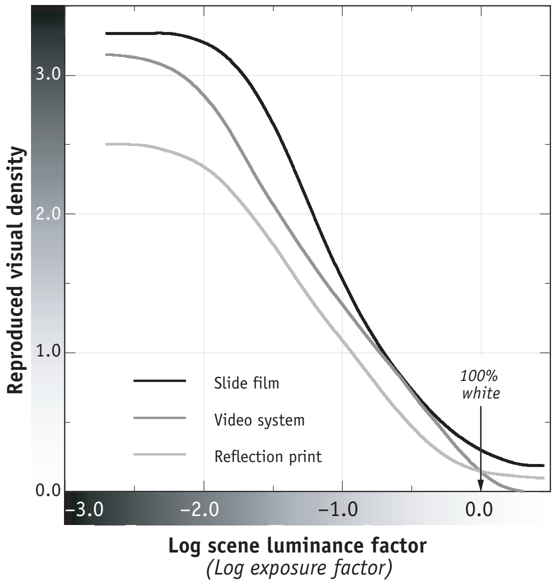


Figure 15.3 Reproduction of a perfect white on three types of media. Each medium can discriminate and represent input information recorded from above 100 % scene white.

also must be designed so that this information can be retained and appropriately represented.

Myth: Output media and devices cannot produce greater than 100 % white; therefore, there is no need to encode information above 100 % white.

Reality: This related color-reproduction myth is a result of confusing inputs and outputs. Consider Figure 15.3, which shows the overall system grayscale characteristics for three types of output: reflection print, photographic slide film, and video display. Note that even the reflection-print medium, which has a minimum density of 0.10 (a reflectance of only about 80 %), still can discriminate and represent *input* information recorded from above 100 % scene white. The "S-shape" characteristic curve compresses the dynamic range of the input information, in both the shadow and highlight regions, such that it can be displayed appropriately. The other two media can discriminate and represent an even greater range of highlight information. Retaining such information is one of the goals for the remaining color-encoding method that will be developed in Part IV.

Discussion

Although reviewing these myths can be somewhat entertaining, the actual consequences of their existence are quite serious. They have caused many color-imaging systems and color-management products to fail, they have caused a great deal of confusion within the color-imaging industry, and they frequently have derailed discussions, which otherwise might have been meaningful and productive, on standards for color interchange.

Readers having remaining questions concerning any of the issues that have been discussed here are urged to review the relevant sections of this book. Some of the concepts that will be discussed in Part IV, although not technically difficult, are fairly subtle and sometimes can be difficult to grasp. Therefore, it is important for the reader to be free of any misconceptions before proceeding.

PART IV

A Unified Color-Management Environment

In the previous parts, the subjects of color measurement, imaging systems, color vision, the colorimetric characteristics of various types of color-imaging devices and media, color-encoding methods, color-encoding data metrics, and output signal processing have been discussed. Each of these topics is an important piece of a much larger picture.

In this final part, that picture will be completed. The individual pieces will be assembled in such a way that a plan is created for a unified color-management environment for the color-imaging industry.

If that plan is implemented, it will be possible for all imaging systems and applications—from the most basic to the most comprehensive—to function together within a single, global, color-managed environment.

Three things are required for the achievement of this objective:

- An overall paradigm that defines how the entire color-managed environment will work.
- A color-encoding scheme that provides appropriate representation of color throughout the environment.
- An overall architecture that allows the environment to be implemented in a practical way.

In the next chapters, the various paradigms that define how color is managed in existing color-imaging systems will be described. It will be shown that although each offers essential (but different) features, no current paradigm is sufficiently extensive in scope to support a truly global color-management

Digital Color Management: Encoding Solutions, Second Edition Edward J. Giorgianni and Thomas E. Madden
 ISBN: 978-0-470-51244-9

environment. To provide that necessary support, a new and comprehensive paradigm we refer to as the *Unified Paradigm* will be described.

In addition, a color-encoding method capable of supporting the color-encoding requirements and color interchange requirements defined by the will be described. This method, which is based on color appearance, was not discussed in Part III. That is because its usefulness is not fully realized outside the context of the comprehensive color-managed environment of the, as will be described in this part.

16

Color-Management Paradigms

Every successful color-imaging system employs one or more means for controlling and adjusting color information throughout the system. That is what is meant by color management. Color management may be incorporated in various forms: as software designed specifically for that purpose, as equipment calibration procedures, as operator adjustments, as chemical process control, etc., used either alone or in various combinations.

At the heart of every color-management approach is an implicitly or explicitly defined *paradigm*—an underlying conceptual model that ultimately determines how an imaging system using that color management will function. One of the problems facing the color-imaging industry today is that there are many different kinds of systems that all seem to work in very different ways. That would suggest there also must be many possible color-management paradigms; if that is so, our goal of a unified, industry-wide, color-managed environment would seem quite hopeless.

A number of years ago, we and a group of our colleagues spent some time thinking about this, and we eventually realized that although existing color-imaging systems might behave quite differently, they all can be described in terms of just *three* fundamental types of color-management paradigms. Each paradigm is perfectly valid, yet each produces very different color results. For convenience, and to avoid any names that unintentionally might imply a value judgment on our part, these paradigms will be referred to simply as Types *A*, *B*, and *C*.

The *ABC* color-management paradigms

Color-imaging systems based on a Type *A* color-management paradigm (Figure 16.1) are "input driven." Their color encoding represents the colors of the input images, and the colors produced by their outputs match (as much as possible) the input-image colors. Color copiers, for example, operate according to the Type *A* paradigm. This generally is the paradigm that first comes to people's minds when they think about color management. In fact, since the paradigm specifies that colors will match throughout an imaging system, the paradigm might seem to be the only one that is needed. But in many ways the basic concept of the paradigm is quite limited. That is why many commercial systems based on Type *B* and Type *C* paradigms also exist.

Systems based on a Type *B* color-management paradigm (Figure 16.2) are "encoding driven." Their color encoding is based on a unifying color-encoding concept that tends to reduce or eliminate the colorimetric differences inherent in the system inputs. For example, some electronic prepress systems encode color in terms of the colorimetric characteristics of a reference graphic arts reflection-print medium.

Digital Color Management: Encoding Solutions, Second Edition Edward J. Giorgianni and Thomas E. Madden
 ISBN: 978-0-470-51244-9

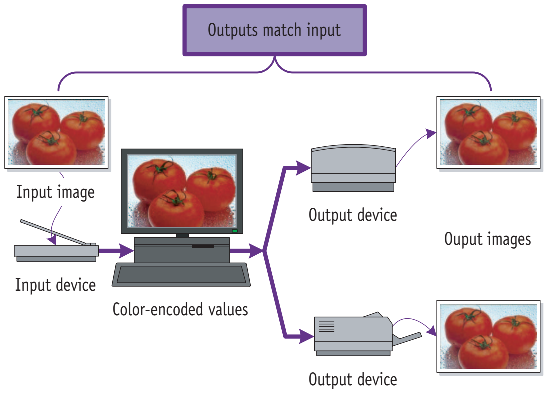


Figure 16.1 Color-imaging systems based on a Type *A* color-management paradigm are "input driven." Their color encoding represents the colors of the input images, and the colors produced by the outputs match the encoded input colors.

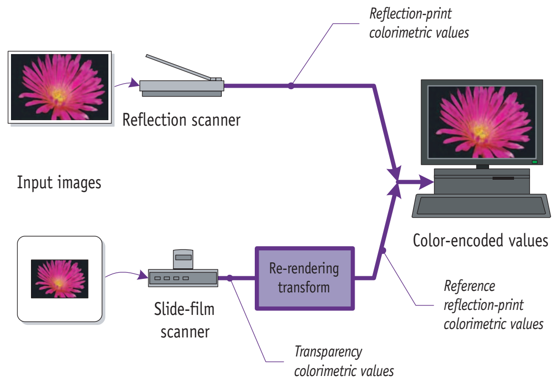


Figure 16.3 An example system, operating according to the Type *B* paradigm. In this electronic prepress system, colorimetric values measured from photographic transparency film inputs are re-rendered to correspond more closely to those of reflection prints.

Colors scanned from actual reflection prints are encoded essentially in terms of their measured colorimetry. But colors scanned from photographic transparency films are *re-rendered*, i.e., their measured colorimetric values are altered such that they correspond more closely to those that typically would be measured from the reference reflection-print medium (Figure 16.3). As in a Type *A* paradigm system, the colors produced by the outputs of a Type *B* paradigm system match the colors represented by the color encoding.

Systems based on a Type *C* color-management paradigm (Figure 16.4) are "output driven." Like Type *B* systems, their color encoding is based on a

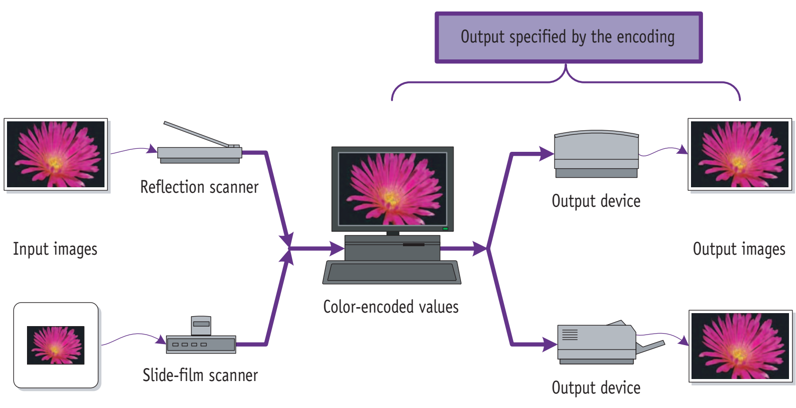


Figure 16.2 Color-imaging systems based on a Type *B* color-management paradigm are "encoding driven." Their color encoding is based on a unifying concept that reduces or eliminates the inherent colorimetric differences of the system inputs. Colors produced by the outputs match the encoded colors.

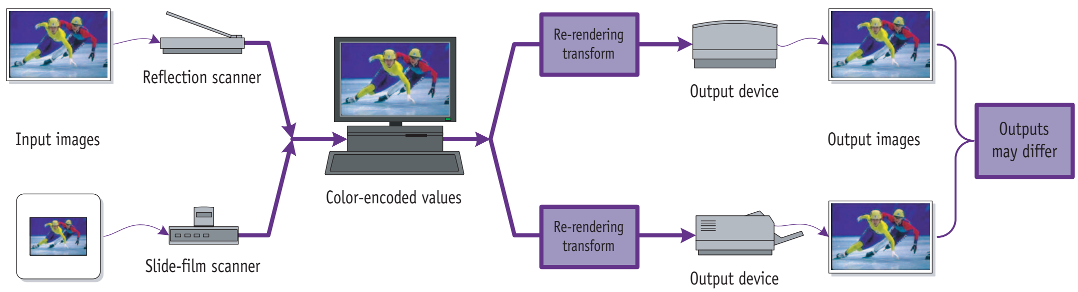


Figure 16.4 Color-imaging systems based on a Type *C* color-management paradigm are "output driven." Images produced on different types of output devices and media will not necessarily match the encoding or each other because various output-specific re-renderings and color enhancements, which may take advantage of the particular capabilities of each output device and/or medium, are performed as part of the output signal processing.

unifying concept. However, their output colors do not necessarily match the colors represented by this encoding because certain *additional* re-rendering is performed, subsequent to encoding, as part of the output signal processing.

The re-rendering performed in a Type *C* system might be done for image simulation, i.e., to make one output produce images that imitate the appearance of images normally produced by another type of output. Output re-rendering also might be employed to enhance the output images by taking advantage of the particular capabilities of each output device or medium. For example, when an output medium having a large color gamut is used, the output signal processing might include some expansion of the gamut of the encoded colors in order to use the full capabilities of that particular medium.

This paradigm often is used in systems where the objective is for each output to produce the best images possible from the encoded data. As a consequence of the output-specific re-renderings and color enhancements that might be performed, images produced on different types of output devices and media will not necessarily match the encoding, nor will they necessarily match each other. This outcome is a principal feature of the paradigm; it is deliberate and should not be thought of as a shortcoming.

Figure 16.5 may help to clarify the distinctions among these paradigms. The figure illustrates the

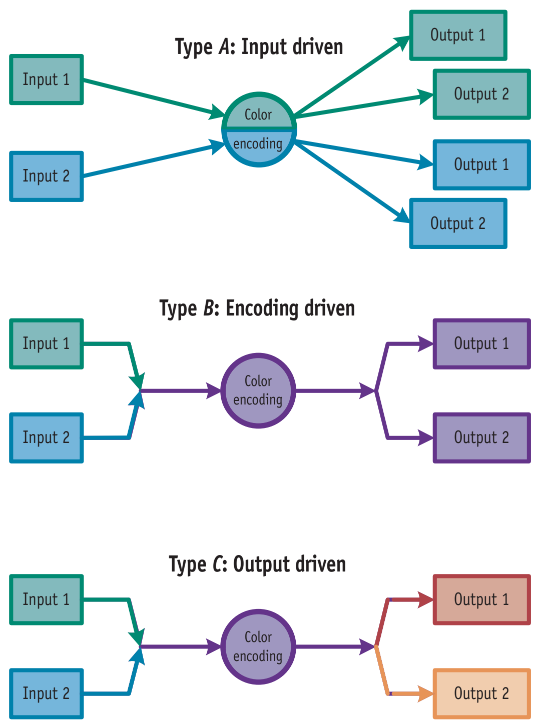


Figure 16.5 Comparison of *ABC* paradigm systems. Each system uses the same two types of input, which differ in color appearance (symbolized by different colors). Each also uses the same two types of output, which also differ from one another. Each system will produce a different color result.

Table 16.1 Comparison of the principal features of Type *A*, Type *B*, and Type *C* color-management paradigms.

Feature	Paradigm *A*	Paradigm *B*	Paradigm *C*
Encoding represents input-image colors	√		
Encoding can represent original-scene colors		√	√
Output colors match encoded colors	√	√	
Seamless compositing of encoded images		√	√
Simpler color-gamut mapping		√	√
Overall system optimization possible			√

different behaviors of systems based on each of the three paradigms. Each system uses the same two types of input—a reflection print medium and a photographic slide film, for example—that differ in color appearance (symbolized by the different colors used in the figure). Each also uses the same two types of output, which also differ from one another.

In a Type *A* paradigm system, the different color appearances of the two inputs are retained and represented in the color encoding; *both* outputs will match, as well as possible, whatever is encoded. So if a reflection print is scanned and encoded, the colors of *both* outputs will match those of that print. If a photographic slide is scanned and encoded, the colors of *both* outputs will match those of that slide. In either case, then, the outputs will match the inputs, and thus the outputs also will match each other (as much as possible).

In a Type *B* paradigm system, color-appearance differences of the two inputs are reduced or eliminated by the color-encoding process. The colors of both outputs will match those represented by the color encoding. Therefore, the outputs also will macolotch each other. However, they will match only one, or perhaps neither, of the inputs, depending on the nature of the unifying concept on which the color encoding is based. For example, if the unifying concept of the color-encoding process is based on the appearance of CMYK reflection prints, images produced on any output device/medium will have the characteristics of CMYK reflection-print images, regardless of the source of the input images and regardless of the characteristics of the actual output device/medium.

In a Type *C* paradigm system, color-appearance differences of the two inputs again are reduced or eliminated by the color-encoding process. However, because output-specific re-renderings and/or color enhancements are performed subsequent to encoding, the outputs (deliberately) will not match the encoded colors, and therefore they may differ from one another.

Feature comparisons

Table 16.1 lists the principal features associated with each of the color-management paradigms. These features are described, and the paradigms are compared, in the following subsections.

Encoding represents input-image colors

Only the Type *A* paradigm uses color encoding that represents the colors of the input image. There are many cases where that representation is the ultimate objective. For example, a photographer may want to create an archival digital record of his or her work. In such cases, the input photographic images would be considered to be *originals*, rather than reproductions, and the digital color encoding should faithfully represent the color appearance of those images.

Encoding can represent original-scene colors

Systems based on either the Type *B* or Type *C* paradigm, but not on the Type *A* paradigm, can be used to encode in terms of original-scene colors rather than in terms of colors as reproduced by the input medium itself. One practical illustration of this important distinction between the Type *A* and Types *B* and *C* paradigms occurs in telecine systems.

Telecine systems are used to scan motion picture films to produce video signals for recording and for television broadcast. Skilled operators, called colorists, make signal processing adjustments during this process in order to produce the best color. Colorists know, however, that the definition of "best" is subject to change, depending on who is making the decision.

For example, if the film being scanned is a theatrical motion picture, the director generally will insist that the television images look like those he or she created on the film. On the other hand, if the film is for a television commercial promoting a product with a trademark color, such as that of a famous soft drink, the director most likely will insist that the video images match the actual color of the product, regardless of whether or not that color was reproduced accurately on the film.

In the hands of a skilled colorist (motivated by a very vocal director) a telecine system effectively can be "switched" between a Type *A* paradigm and a Type *B* or Type *C* paradigm. There is a very important lesson here that will be revisited in Chapter 17.

Output colors match encoded colors

In the Type *A* and Type *B* paradigms, but not in the Type *C* paradigm, the system outputs match the color specified by the encoding (as closely as possible, given the limitations of the output color gamuts). This matching would be an important feature if, for example, a system were used to produce output color samples according to colorimetry specified in terms of the color encoding, or if it were used to produce output from a computer-generated image that was defined in terms of color-encoded values.

Seamless compositing of encoded images

The unique color characteristics of different devices and media—the very characteristics that manufacturers work to build into their products to differentiate them—are regarded mostly as a nuisance in applications where portions of images from different inputs are merged to form composite images. When these color characteristics are retained, as they are in the Type *A* paradigm, the boundaries of the merged image areas can be very apparent, depending on the disparity of the input sources. In the Type *B* and Type *C* paradigms, color differences among the inputs are minimized, which results in more homogeneous-appearing composites.

Simpler color-gamut mapping

Gamut mapping is somewhat simpler in systems based on either the Type *B* or Type *C* paradigms than in systems based on the Type *A* paradigm. Mapping between two color gamuts—in this case, the gamut of the encoded colors and the gamut of the output colors—is more straightforward when both are well defined. Because the mapping is performed as part of the output signal processing of each individual output device/medium, the output color gamut can be explicitly defined. By comparison, in the Type *A* paradigm, the gamut of encoded colors will vary, depending on the color gamuts of the particular media that have been used for input. This makes it somewhat more difficult (but certainly not impossible) to develop a single, input-independent gamut-mapping transform for each output. Because the color-encoding methods used in the Type *B* and Type *C* paradigm systems minimize color differences among the inputs, the potential color gamut of the encoded values is more consistent and therefore more easily defined.

Overall system optimization

Only the Type *C* paradigm allows optimization of the color reproduction of an entire imaging system, from original scene to final output image. In a Type *A* paradigm system, the output image can be no better than the input image (by definition, it cannot be better if it matches). An imaging system based on the Type *B* paradigm can do better, in that a color-encoding method that extracts original-scene colorimetry from reproduced colors can be incorporated. But since, in the Type *B* paradigm, a further objective is for the outputs to match each other, optimum use might not be made of the capabilities unique to each particular output. In a Type *C* paradigm system, the objective instead could be for each output to produce the best possible rendition of an encoded image. If the output media are different—say a video display, a reflection print, and a photographic slide—the best renditions certainly will differ somewhat from medium to medium.

Discussion

These three paradigms appear to be sufficient for describing the basic functionality of all existing types of color-managed imaging systems. That is quite encouraging—or at least certainly more encouraging than if *dozens* of different paradigms were required. Yet to truly unite the color-imaging industry, and to achieve the goal of unrestricted color interchange among systems, it would seem necessary for all systems to operate from a *single* paradigm. But which of the three should it be? On what basis could a selection be made?

It cannot be made based on functionality. The color encoding and output results of systems based on each paradigm are different. Each paradigm has features that are essential to large numbers of users, and none of the three has all the features of the others. Nor can a selection be made based on technical merits. Successful color-imaging systems have been built based on each of the paradigms. Each is technically sound and practical to implement.

Numerous attempts have been made to negotiate a selection by industry-wide agreement. (Although we never heard the concept of an "underlying paradigm" expressed during these negotiations, in retrospect it is now apparent that various paradigms indeed were being proposed and discussed.) These attempts have failed because, we believe, all of the proposed paradigms have been too limited. The participants in such negotiations represent profit-making corporations. Understandably, they have been reluctant to give up system features that they know are important to their customers. As a result, the industry has continued to work with numerous alternative—and essentially incompatible—types of systems.

We, and a number of our colleagues, felt strongly that the color-imaging industry needed a *complete* solution to color management—one that would eliminate such incompatibilities. This required the development of an overall system architecture based on a single, comprehensive, unified color-management paradigm. That paradigm is the topic of Chapter 17.

Summary of key issues

- All successful color-imaging systems employ some form of color management for controlling and adjusting color information throughout the system.
- Any color-management approach is based, implicitly or explicitly, on an underlying paradigm that defines how color-imaging systems using that color management will behave.
- All existing color-imaging systems can be described in terms of three fundamental types of color-management paradigms.
- In the Type *A* paradigm, color encoding is based on the colors of the input image, and output colors match the input colors.
- In the Type *B* paradigm, color encoding is based on a unifying concept, such as a reference input or output. Output colors match the colors represented by that encoding.
- In the Type *C* paradigm, color encoding again is based on a unifying concept. However, the output colors generally differ from those represented by the encoding due to deliberate color enhancements and re-renderings performed subsequent to encoding.
- Each paradigm has features that are essential for various applications, and none has all the features of the other two.
- A complete solution to color management and color interchange requires a unified color-management paradigm.

17

A Unified Paradigm: Basic Properties

If the hope of unification should prove well founded, how great and mighty and sublime. . .
Michael Faraday

At the time we were wrestling with the problems resulting from the use of three different color-management paradigms within the color-imaging industry, we both happened to be reading a book entitled *Hyperspace*. In that book, theoretical physicist Michio Kaku describes a problem that scientists puzzled over for decades: why do the fundamental forces of the universe—gravity, electromagnetism, and the strong and weak nuclear forces—require markedly different mathematical descriptions?

Kaku explains that these apparent differences appear only because the basic paradigms used in their description are too limited. He shows that if all the fundamental forces are seen according to a larger *unified* paradigm—as vibrations in higher-dimensional spaces—their field equations suddenly unite in what he describes as an "elegant and astonishingly simple form."

Now, developing a color-management system is hardly the equivalent of trying to explain the workings of the entire universe (although there are times when we are not sure which would be easier). Nevertheless, there are parallels, and there are lessons that can be learned from the search for a unified field theory.

Perhaps the most important lesson is that, historically, when true understanding is reached, the results exhibit unity, simplicity, beauty, and elegance. For example, the eight Maxwell equations for electricity and magnetism are notoriously ugly, because time and space are treated separately. But when rewritten in relativistic form, where time is treated as a fourth dimension, the separate equations collapse to a single equation that is simple and elegant in form. Similarly, centuries ago, when the paradigm for the solar system placed the earth at its center, describing the motions of the planets required complex functions. When the system became better understood and could be envisioned in a larger paradigm, the motions were shown to be simple ellipses.

In a sense, the technology of color management seemed to be at the "pre-understanding" stage. The need to use three different paradigms to explain the color-imaging "universe" certainly indicated a lack of "unity and simplicity," and the methods required to transfer images and control color across systems were

Digital Color Management: Encoding Solutions, Second Edition Edward J. Giorgianni and Thomas E. Madden
 ISBN: 978-0-470-51244-9

anything but "beautiful and elegant." But perhaps, we thought, if the *ABC* paradigms could be seen according to some larger paradigm, they also would be found to fit together in an "elegant and astonishingly simple form." Moreover, if a single paradigm uniting the *ABC* paradigms could be found, the first requirement for building a unified color-management environment for the industry would be met.

An ideal application

To determine the basic requirements of this new paradigm, we and our colleagues began by imagining an "ideal application." We tried to envision a color-management application, such as a desktop imaging program, that was capable of making an imaging system perform in *all* the ways that users of any system, based on any of the three *ABC* paradigms, might want. We were not necessarily proposing that such an application be developed. The concept of an ideal application was used only to help focus on what color-managed systems should *do*, rather than on the technical details of how they might be designed and implemented. That would come later.

So what *should* color-managed systems do? That turned out to be a surprisingly difficult question to answer. For example, consider the system shown in Figure 17.1. The figure shows an original object (a London bus) being photographed to produce a 35 mm slide. The slide is scanned to produce a digital image file on the computer workstation. That image file then is used to produce another 35 mm slide from a photographic film writer, an image on the workstation monitor, and a reflection print from a thermal printer. (Note that we have used a photographic slide film in this example and elsewhere in this part because slide films can function as both input and output media on digital color-imaging systems. As such, their inclusion helps to demonstrate how various desired input/output relationships can be accommodated in the color-management paradigm being discussed.)

If this example system were color managed, what should its images look like? When a number of people were asked that question, many different answers were obtained. But perhaps that should have been expected. There are many possible answers because there are many reasons why color images are produced.

Let us look at how a single factor can generate a whole series of consequences. Assume that the slide film reproduces reds as somewhat orange, as some films tend to do. Now, should the system produce an

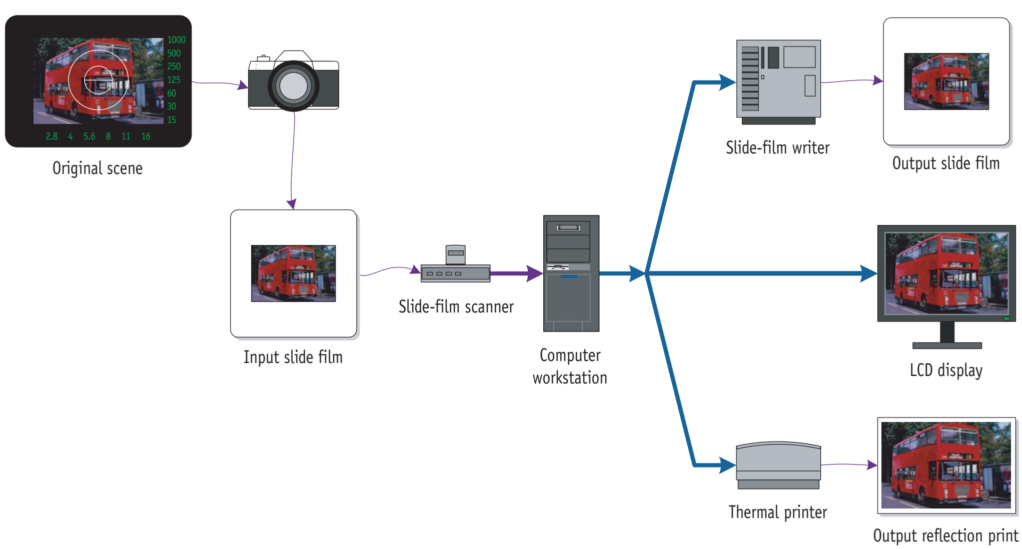


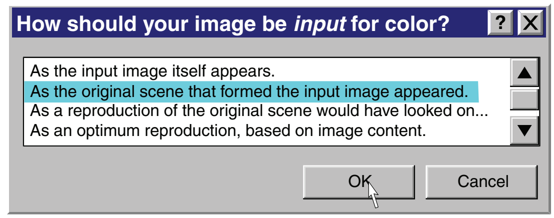
Figure 17.1 What should this color-managed system do?

output slide with an orange-red bus, thus matching the input slide, or should it alter the input color values such that the output red matches the original red of the bus? What should the monitor image look like? Should it match the input slide or the altered image that more accurately represents the original color of the bus? And what about the thermal print? Should it match the output slide? The input slide? The monitor image?

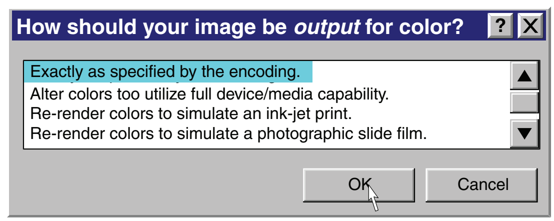
As was discussed earlier, there are no right or wrong answers to these kinds of questions. The "right" answers will vary depending on the purpose for which a system is used. In fact, in most current systems, such questions are never explicitly posed to the user. The answers instead are "hard coded" by the system designers on the basis of certain assumptions they have made about the results that system users will want.

Although a fixed, predetermined behavior may be perfectly appropriate for an individual system, an ideal application must provide much greater flexibility. It must allow the user to alter the *entire system behavior*—in much the same way that was described earlier in the example of the telecine system—in order to produce different output results. As we considered the telecine system and a wide variety of other types of imaging systems, it became apparent that in order to provide functionality that is sufficiently flexible to cover *all* potential uses, an ideal application would have to ask the user three basic questions concerning input, output, and interchange.

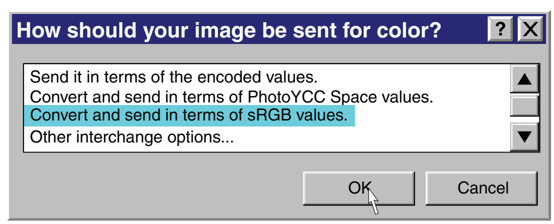
The first question would ask the user to define what interpretation will be given to an image that is about to be input to the system. The image could be interpreted as an original, in which case the user would want to encode the image in terms of its own appearance under some specified set of viewing conditions. The image instead could be interpreted as a record of an original scene, in which case the user would want to encode the image in terms of the appearance of that scene. Another interpretation might be that the image is considered simply as a source of information from which various renderings can be made. For example, the image could be encoded as if it had originated on another medium, or it might be re-rendered according to any number of different criteria. These and perhaps other different input interpretations can be specified by responding to the following query:



The next question would ask the user to define an interpretation for images that are to be output by the system. Although there are many possible variations, the responses generally would fall into the principal categories shown below.



The final question would ask the user to define how encoded images will be transferred to another imaging system.



A color-imaging system controlled using an application that included these questions could be altered to duplicate the functionality of *any* of the *ABC* color-management paradigms. For example, if one were to scan a photographic slide and *answer* "As the input image itself appears (under a specified set of viewing conditions)" for input and "Exactly as specified by the input option" for output to a reflection thermal printer, the system would produce a reflection print that matches, as much as possible, the appearance of the input slide. This is a Type A paradigm result, as shown in Figure 17.2.

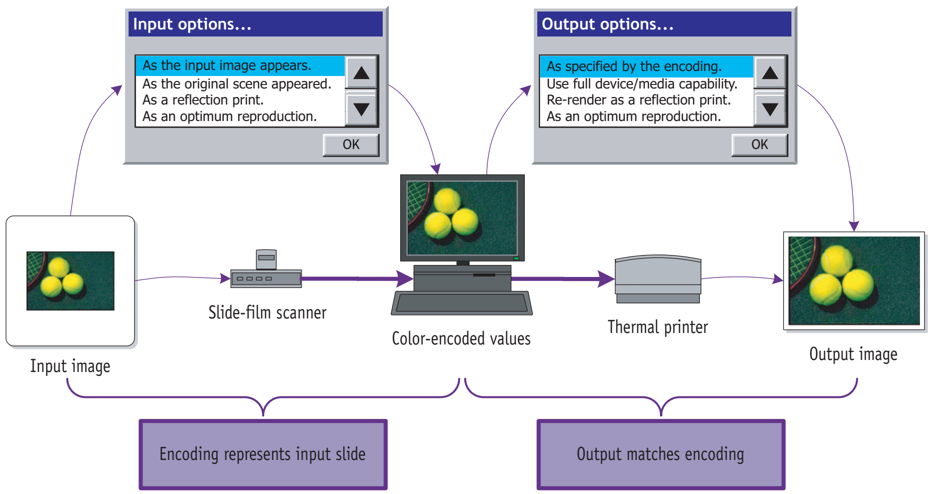


Figure 17.2 In this example, the ideal application produces a Type *A* paradigm result. The appearance of the output reflection print matches that of the input slide.

If one were to again scan a photographic slide, but instead answer "As a reproduction of the original scene would have looked on (a particular reflection-print system)" for input, and again answer "Exactly as specified by the input option" for output to a reflection thermal printer, the system would produce a thermal print that matches the appearance of a print that would have been produced by the specified reflection-print system. Because the encoding and output print both represent a re-rendering of the input slide, rather than the appearance of the slide itself, this is a Type B paradigm result (Figure 17.3).

If one were to scan a reflection print and answer "As the input image itself appears (under some specified set of viewing conditions) " for input, and "Alter the color specified by the input option; use the full capabilities of the output" for output to a photographic slide-film writer, the system would produce a Type *C* paradigm result, as shown in Figure 17.4. Because the color gamut of the output slide film is somewhat greater than that of the input reflection-print medium, the gamut of the encoded colors would be expanded such that the gamut capabilities of the slide film are best utilized.

These examples demonstrate that a single system, controlled from a sufficiently flexible application, can be made to behave according to any of the *ABC* paradigms. The system can be directed such that the outputs match the input (the Type *A* paradigm). Alternatively, it can be directed such that the encoding is based on the properties of reference input or output and the outputs match that encoding (the Type B paradigm). As a third alternative, the system can be directed such that final outcome is determined by the specific characteristics of each output device/medium (the Type *C* paradigm).

From an ideal application to a unified paradigm

The fact that the different behaviors of the *ABC* paradigms can be emulated using a single application supports the premise that all existing color-imaging systems are subsets of a single larger paradigm, which we refer to as the Unified Paradigm. Moreover, because that paradigm must be capable of *supporting* the ideal application, the features of that application essentially define the properties of the paradigm

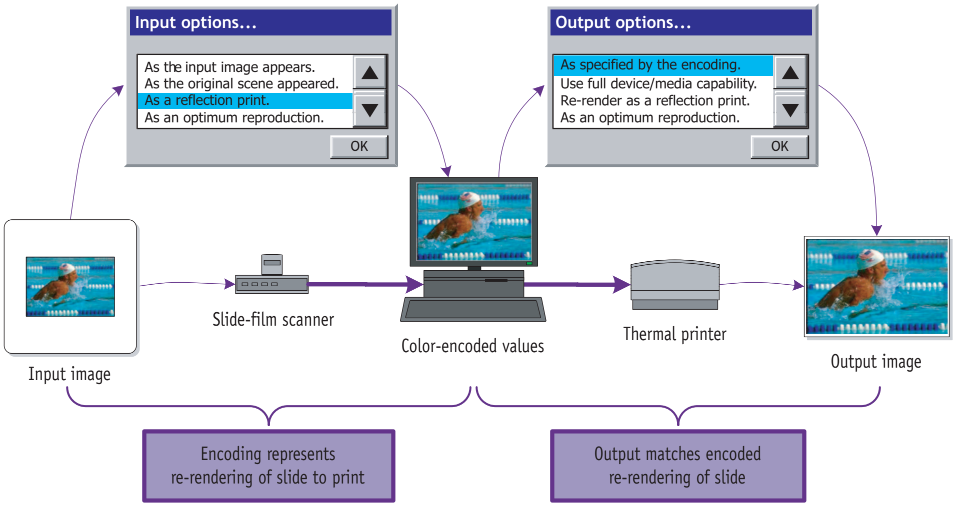


Figure 17.3 In this example, the ideal application produces a Type *B* paradigm result. The appearance of the output thermal print matches that of a print that would have been produced by the reflection-print system specified in the encoding option.

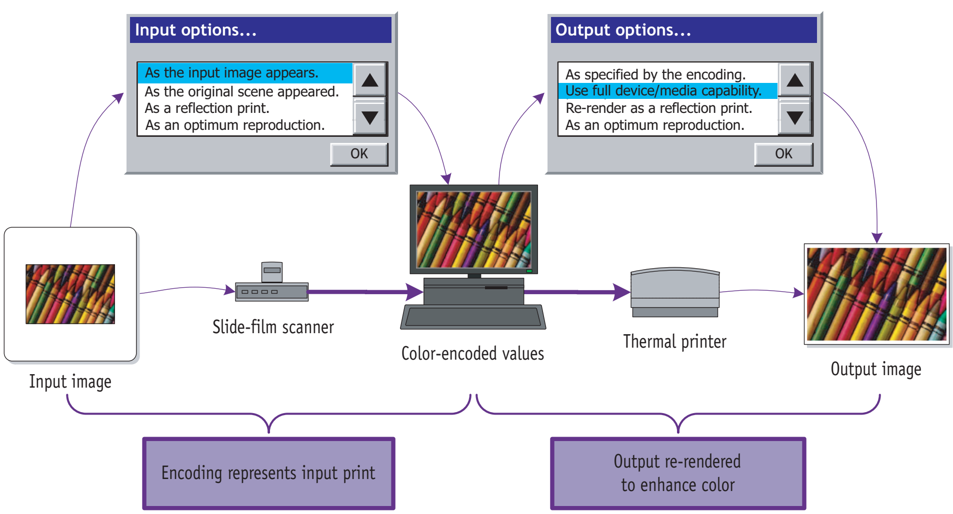


Figure 17.4 In this example, the ideal application produces a Type *C* paradigm result. The colors represented by the encoding are enhanced for output to the slide film.

itself. Specifically, the Unified Paradigm must do the following:

- Support all types of input and output devices and media.
- Provide for multiple input interpretations, which must not be predetermined or limited by the paradigm itself.
- Provide for multiple output interpretations, which also must not be predetermined or limited by the paradigm.
- Allow input and output interpretation decisions to be made *independently* of one another; all possible input and output interpretation combinations must be allowed.

In addition, the paradigm must support the interchange of images among systems within a unified color-management environment. As was discussed earlier, different color data metrics may be most appropriate for the internal encoding of different systems. For that reason, the Unified Paradigm also must do the following:

- Support the use of various internal and interchange color-encoding data metrics.
- Allow the selection of interchange color-encoding data metrics to be made independently of input and output interpretation selections.

These basic properties of the Unified Paradigm are shown in diagrammatic form in Figure 17.5.

Discussion

In this chapter we have discussed, enumerated, and illustrated the basic requirements of a unified color-management paradigm. But can a practical color-managed environment actually be *built* according to a paradigm incorporating those requirements? After all, the Unified Paradigm supports every form of input and output, it allows the freedom to choose from an unlimited selection of input and output interpretations, it places no restrictions on the combinations of inputs, outputs, and interpretations that can be selected, and it supports all color-interchange data metrics. Is that really a meaningful color-management paradigm, or is it simply a prescription for anarchy?

To some, it might seem that in order to prevent a color-managed environment based on such an unrestricted paradigm from falling into complete chaos, at least some of the Unified Paradigm's flexibility would have to be constrained or even eliminated in actual practice. That is, in effect, what most color-management system architects have concluded. As a result, available color-management systems generally are based on just one of the constituent paradigms supported by the Unified Paradigm. Most often, that is the Type *B* paradigm. The principal reason for this is that input compatibility and system-to-system interoperability are easily achieved under that paradigm because all input images are rendered or re-rendered as necessary such that they conform to a common appearance.

As discussed in Chapter 16, such Type *B* paradigm systems have features that make them appropriate for certain applications, particularly those that produce images primarily for a single type of output. Like their Type *A* and Type *C* counterparts, however, Type *B* systems also lack features and capabilities that often are essential for other types of applications. When color management is based solely on any one of these limited paradigms, some types of imaging-system behaviors become unavailable. The overall industry's needs for those behaviors, however, will always be present. Color-management systems must meet diverse customer requirements; those requirements cannot simply be ignored or altered to conform to the limitations of particular systems. As experience has shown, and as would be expected, practitioners will bypass the rules of restricted-paradigm systems—or avoid them entirely—when that is necessary in order to achieve particular required results. The inevitable outcome of such necessary measures is the creation of even greater disparities and incompatibilities among various types of color-imaging systems.

Our position is that the retention of *all* the capabilities of the Unified Paradigm is essential in order to fully meet the collective needs of the imaging industry. We also believe that those capabilities can be implemented in practical systems, without restrictions and without limitations. A fundamental

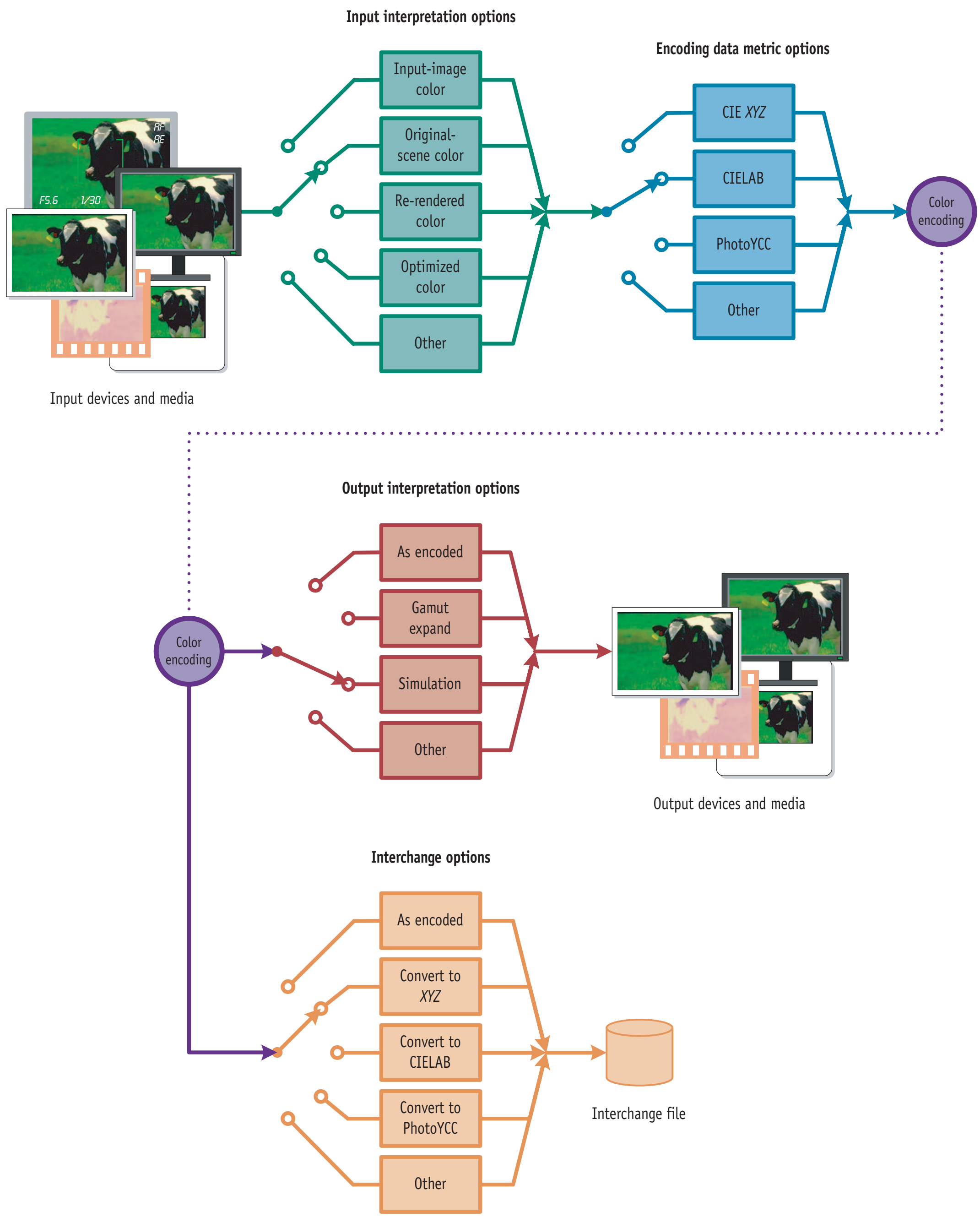


Figure 17.5 A diagram of the Unified Paradigm.

requirement for the successful implementation of the Unified Paradigm is the use of an appropriate method for encoding color throughout the color-managed environment. That method is the topic of the next chapter.

Summary of key issues

- The different behaviors of the Types *A*, *B*, and *C* color-management paradigms can be emulated by a single, unified paradigm.
- The Unified Paradigm supports all types of input and output devices and media, it allows multiple input and output interpretations, and it supports multiple interchange data metrics.
- In the Unified Paradigm, input and output interpretation decisions are independent of one another, and all possible input/output interpretation combinations are allowed.
- In the Unified Paradigm, the selections of color-encoding data metrics and color-interchange data metrics are independent of input and output interpretation decisions.
- The successful implementation of the unified color-management paradigm requires an appropriate method for encoding color.

18
A Unified Paradigm: Encoding Concepts

The Unified Paradigm introduced in the preceding chapter offers the promise of a comprehensive, inclusive, and unified color-management environment. However, realization of that promise requires the use of a color-encoding method that is consistent with, and fully supportive of, all aspects of the paradigm.

In this chapter, we will further explore the Unified Paradigm and introduce the basic concepts of a color-encoding method that enables the complete potential of the paradigm to be realized. The colorimetric transformations required to implement the encoding will be described in Chapters 19. How the color-encoding method can be used in practice then will be demonstrated in Chapter 20 and 21, where the Unified Paradigm, color-encoding concepts, and colorimetric transformations will be applied to a number of example color-imaging systems.

Color-encoding requirements

Figure 18.1 illustrates the application of the Unified Paradigm to a color-imaging system having a single input and a single output. This system, operating on its own, is capable of performing all the functions of the "ideal application" that was used earlier to help define the properties of the paradigm. Although the basic signal processing arrangement of this system is the same as that used in previous examples, the multiple input/output interpretation options of the paradigm have been incorporated. Each option is implemented in the form of an individual signal processing transform. In Figure 18.2, the paradigm has been applied similarly to a more complex multiple input/output system.

For systems such as these to work successfully, the color-encoding method they use must, of course, achieve compatibility among all system inputs. However, the *definition* of what constitutes an "input" must be expanded.

In the Unified Paradigm, an input no longer is just a particular medium; it is a particular *interpretation* of the color from a particular medium. *Each interpretation from each medium constitutes a unique input source*. What this means is that the color-encoding method of the Unified Paradigm must not only achieve input compatibility, but do so *without implicitly assigning its own particular interpretation* on encoded colors.

None of the color-encoding methods described so far meets these requirements. Some fail because they cannot achieve compatibility among all input media, while others fail because they do not allow multiple interpretations of color. For example, in Part III it was shown that densitometric and standard colorimetric encoding methods do not achieve

Digital Color Management: Encoding Solutions, Second Edition Edward J. Giorgianni and Thomas E. Madden
 ISBN: 978-0-470-51244-9

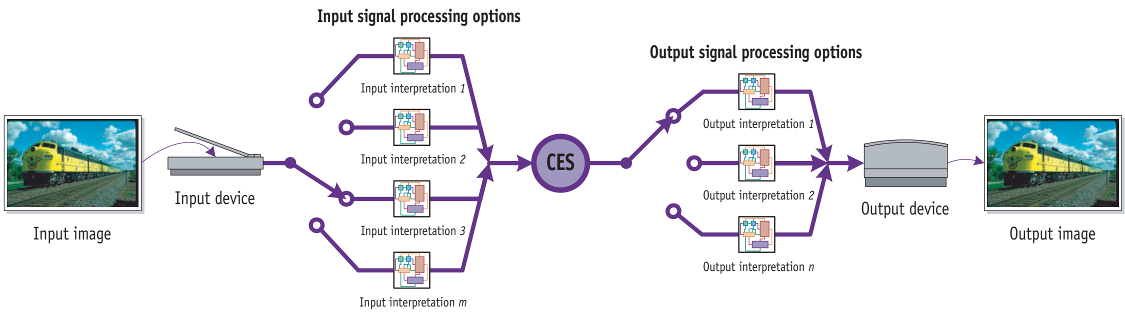


Figure 18.1 Application of the Unified Paradigm to a single-input imaging system. The incorporation of multiple input/output interpretation options makes this system capable of performing all functions defined by the paradigm.

compatibility among disparate types of input media. Use of those methods therefore restricts the types of input media and devices that can be supported.

On the other hand, the scene-based encoding method used on most Photo CD imaging workstations, the rendered-print encoding method used by many graphic arts prepress systems, and other Type *B* or Type *C* paradigm encoding methods can achieve input compatibility among disparate forms of input media. However, they do so by imposing particular interpretations of input color. Those interpretations subsequently restrict the behaviors of systems on which the encoding methods are used.

How can this apparent conflict between unrestricted functionality and system interoperability be resolved? Throughout this book, each new color-encoding problem has been solved successfully by basing the solution on the principle of input compatibility. In each new situation, a specific aspect of color that all system inputs had in common was identified, and an appropriate color-encoding method was selected or developed based on that color aspect.

From a retrospective examination of those problems, it can be seen that their solutions have followed a definite pattern: as the types of inputs became more and more disparate, the common color aspect itself had to be more and more inclusive. In a sense, the logical conclusion of that process now has been reached. Since the Unified Paradigm must support multiple input types and multiple input color interpretations, *the only thing the inputs now have in common is color itself*.

This means that in order to realize the promise of the Unified Paradigm, it will be necessary to answer a question that was raised at the start of this book: "How can *color* be represented in numerical form?"

Representing color

Of the encoding methods that have been discussed so far, those based on CIE colorimetry come the closest to representing color (i.e., color *appearance*). But as has been shown, standard CIE colorimetry alone does not represent color appearance unambiguously. The same color stimulus defined by a set of CIE *XYZ* tristimulus or CIELAB L^*, a^*, b^* values can have any number of entirely different appearances, depending on the conditions under which the stimulus is viewed. That ambiguity is a serious problem, because there are three different types of viewing environments involved in the color-imaging process (Figure 18.3), and there are limitless possible sets of viewing conditions for each of those types.

On the input side of an imaging system, there are *original-scene environments*, i.e., the environments in which live original scenes are viewed and captured. Also, on the input side, there are *input-image environments*, where hardcopy and softcopy images that are to be input to a color-imaging system are viewed. Finally, there are *output-image environments*, where hardcopy and softcopy images produced by a color-imaging system ultimately are viewed.

One way to deal with the complications caused by the different viewing conditions of these environments would be to impose a restriction that

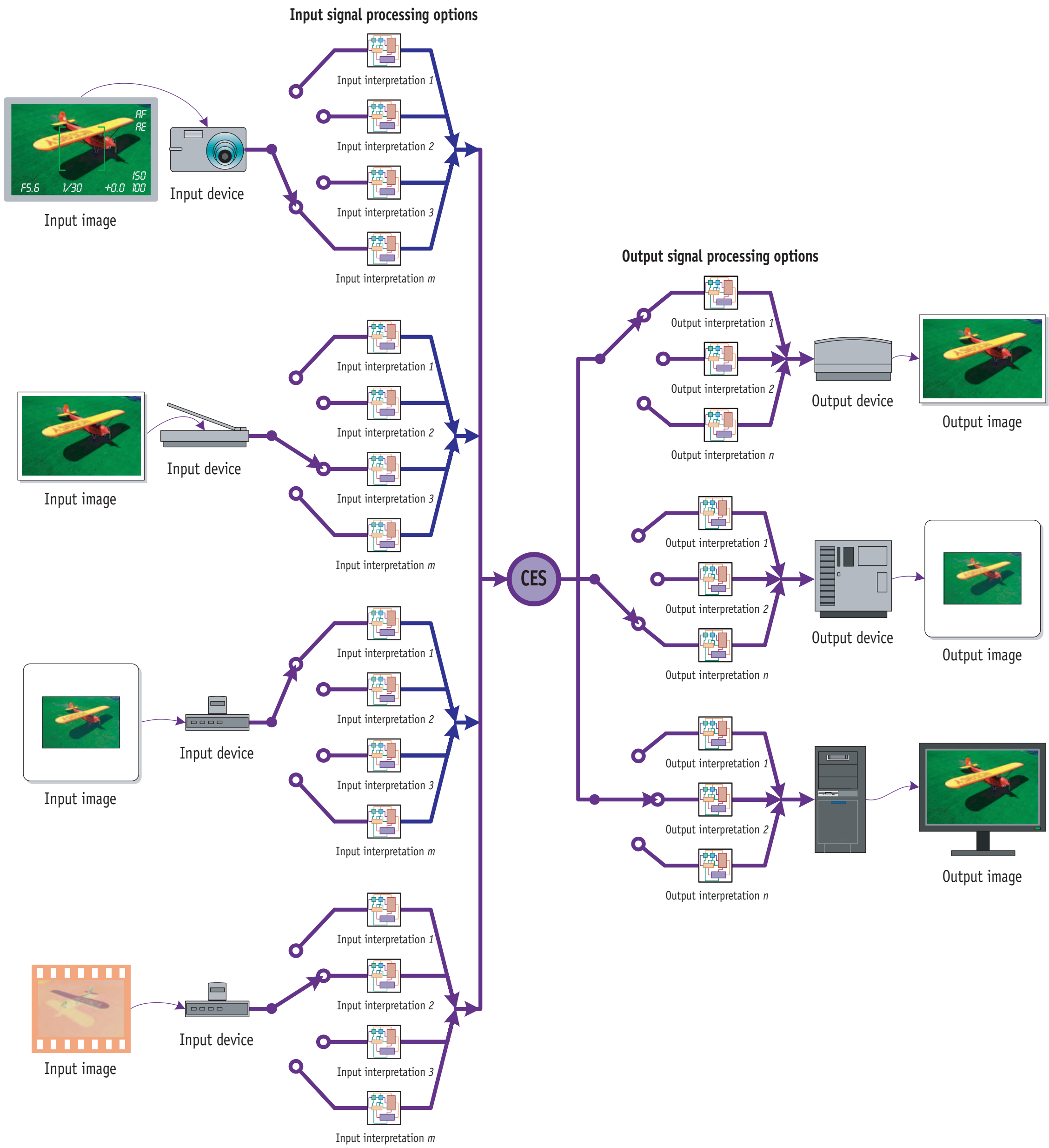


Figure 18.2 Application of the Unified Paradigm to a more complex imaging system.

an *identical* set of conditions must be used for original-scene viewing and image capture, for input-image viewing, and for output-image viewing. In that case, a colorimetric specification alone would be sufficient to represent color. The color-encoding methods of some existing systems are, in fact, based on that type of restriction. For example, the manufacturer of a reflection copy system may specify that an output image produced by the system will be an accurate copy of an input image, but only if both

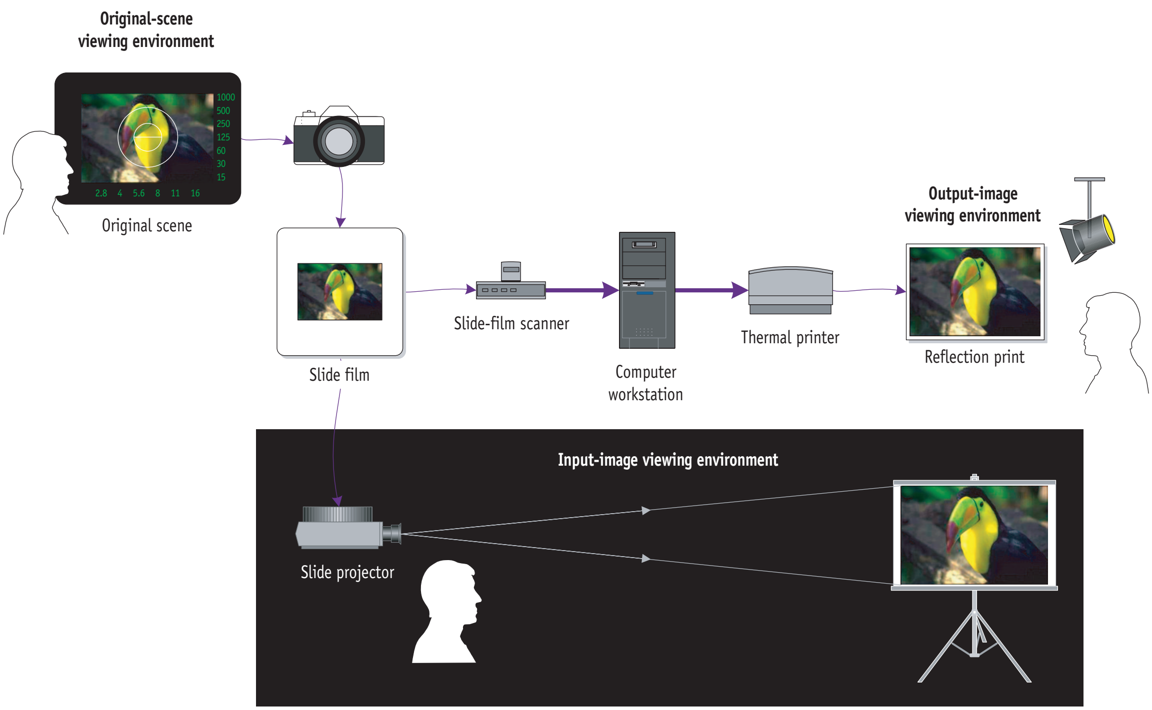


Figure 18.3 There are three basic types of viewing environments associated with the color-imaging process: original-scene, input-image, and output-image environments. The conditions of each can differ greatly.

images are viewed in a specified graphic arts viewing booth.

Placing any such restrictions would be inappropriate for our intended purpose. Instead, what is needed is a color-encoding method capable of providing *all* the features of the Unified Paradigm. That method therefore must do the following:

- Allow live scenes to be viewed and captured under any original-scene conditions.
- Allow input images to be viewed under any desired conditions and to be encoded in terms of their appearance under those particular input-image viewing conditions.
- Allow output images to be tailored specifically for display in any desired set of output-image viewing conditions.

Ironically, the real solution to this problem is not to *restrict* the conditions of the original-scene, input-image, or output-image viewing environments. Quite the contrary, the solution is to add yet *another* viewing environment with its own set of conditions.

Encoding reference viewing conditions

The viewing conditions for this additional environment will be referred to as *encoding reference viewing conditions*. The following hypothetical examples (Figures 18.4a and 18.4b) illustrate how the use of these reference conditions can help to represent color unambiguously.

Suppose one person (the sender) wanted to communicate a color to another person (the receiver), and suppose that they each had identical devices that could generate any color stimulus that was specified in terms of standard CIE colorimetric values. If the sender communicated a set of such values for a color, they could be input to the receiver's device.

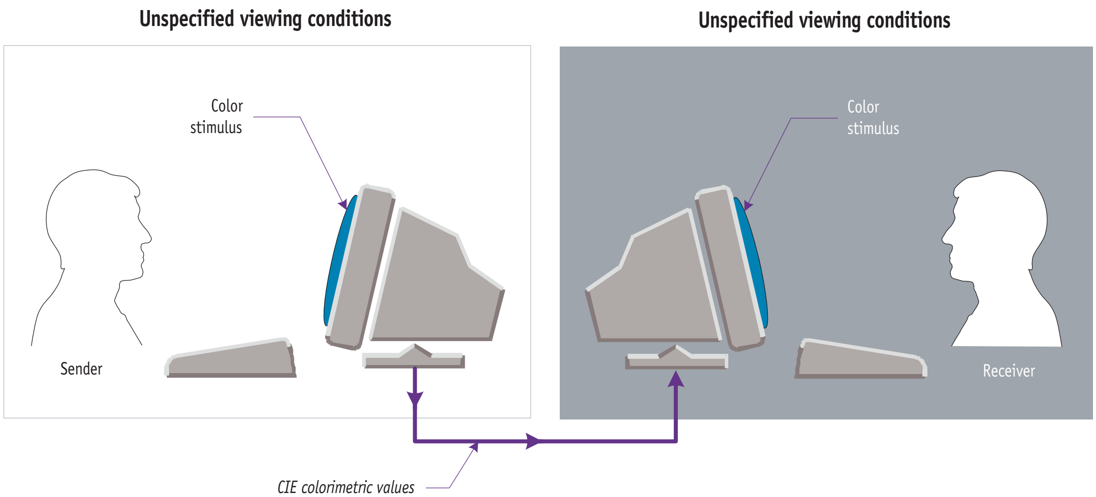


Figure 18.4a Color representation and communication based on CIE colorimetric values alone is ambiguous.

That device then would generate the color *stimulus* the sender intended. But it would not be certain whether the right *color* had been communicated or not, because the appearance of the stimulus would change depending on the conditions in which it was viewed.

That problem can be solved if the sender and receiver agree on a set of reference viewing conditions (Figure 18.4b). The sender always would view the color-stimulus device under those conditions and would adjust the colorimetric values until the device generated a stimulus having the color that he or she wants to communicate. Those values would be sent to the stimulus generator of the receiver, and the receiver would view the resulting stimulus under an identical set of viewing conditions. The receiver

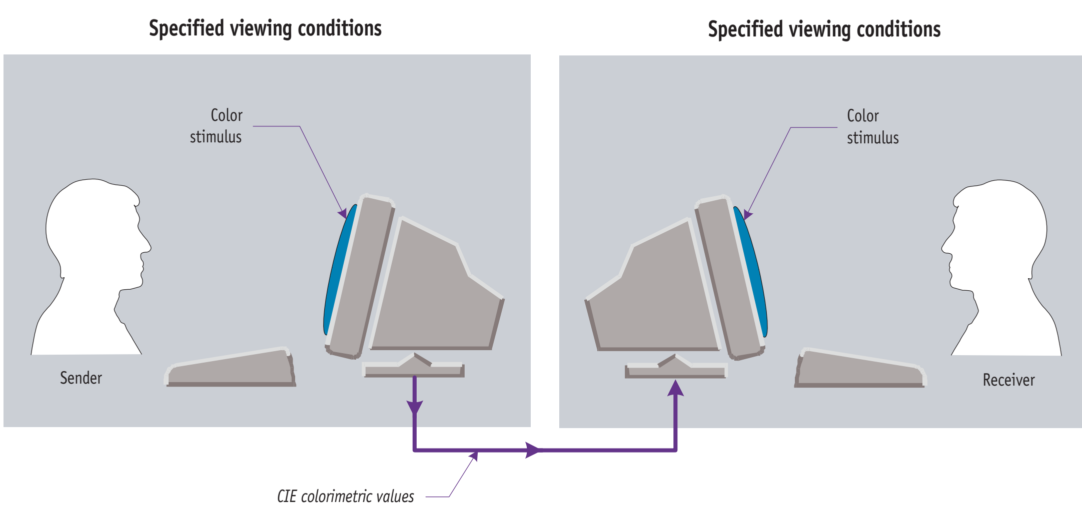


Figure 18.4b Color representation and communication based on CIE colorimetric values and a set of reference viewing conditions is unambiguous.

then would see the *color* the sender intended to communicate (assuming, of course, that the sender and receiver have identical color vision).

A color-encoding method capable of supporting all the objectives of the Unified Paradigm can be based on this strategy for color communication. Encoded colors can be expressed in terms of colorimetric values, but the ambiguity that is inherent in a colorimetric specification alone is eliminated by the additional specification of a set of encoding reference viewing conditions. The concept of this encoding method can be stated as follows:

> *In the unified color-management paradigm, encoded values specify the colorimetry for a stimulus that, when viewed according to a specified set of reference conditions, would produce an intended color appearance.*

This encoding method essentially provides a "recipe" for color representation. The combination of a colorimetric specification and a defined set of viewing conditions provides a complete formula for producing and displaying a color stimulus having the intended color appearance.

Colors encoded according to this method are *unambiguous*, because there is one and only one colorimetric specification that can produce a particular color appearance (to a standard observer) under the specified viewing conditions. At the same time, the method does not impose a particular interpretation on encoded colors. The "intended color appearance" can refer to an original-scene color, a color of a viewed input image, a user-created or user-edited color on a computer monitor, a color to be output for viewing, or any other type of color that a user wishes to represent.

It is important to emphasize that the use of a set of encoding reference conditions in no way restricts the behavior or flexibility of an individual system operating according to the Unified Paradigm. That is because the reference conditions do *not* specify the conditions that must be used for viewing and capturing original scenes, nor do they specify the input-image or output-image viewing conditions. Any light source spectral power distribution, any level of illumination, any amount of viewing flare, and any surround conditions can be used for the capture and display of actual system images.

In fact, no one need ever construct or use the reference conditions for actual viewing. The encoding reference viewing conditions apply *only to the color-encoding process itself.*

Color-encoding guidelines

Figure 18.5 illustrates the overall concept of the color-encoding process. The color-encoding method is such that colors are represented in a color encoding specification (CES) in terms of their CIE colorimetric values, as measured in the absence of flare light. The CES includes a description of the encoding reference viewing conditions. Encoded values are expressed numerically in terms of the particular color-encoding data metric defined for the CES. As discussed earlier, the selection of a data metric for a given system is an engineering decision, so any of a number of different data metrics can be used.

As the figure shows, input signal processing transforms are used to convert the colorimetry of color stimuli in actual image-capture and input-image viewing conditions to corresponding colorimetry for equivalent stimuli viewed in the encoding reference conditions. Similarly, output signal processing transforms are used to convert encoded colorimetric values to corresponding values for equivalent stimuli to be viewed in the actual output-image viewing conditions. In special cases where the actual input or output viewing conditions are *identical* to the encoding reference viewing conditions, such conversions are not required. More typically, colorimetric transformations are needed in order to account for *differences* in the actual and encoding reference conditions.

The colorimetric transformations used in this appearance-based color encoding must first account for any physical alterations that the various viewing conditions might produce on an image. The transformations also must account for any alterations that the conditions might induce in an observer's perception of color. While accounting for physical factors (primarily differences in the amount and/or chromaticity of flare light) is quite straightforward, accounting for effects resulting from perceptual phenomena is somewhat more complex.

Over the years, there has been a great deal of theoretical and experimental work done to better

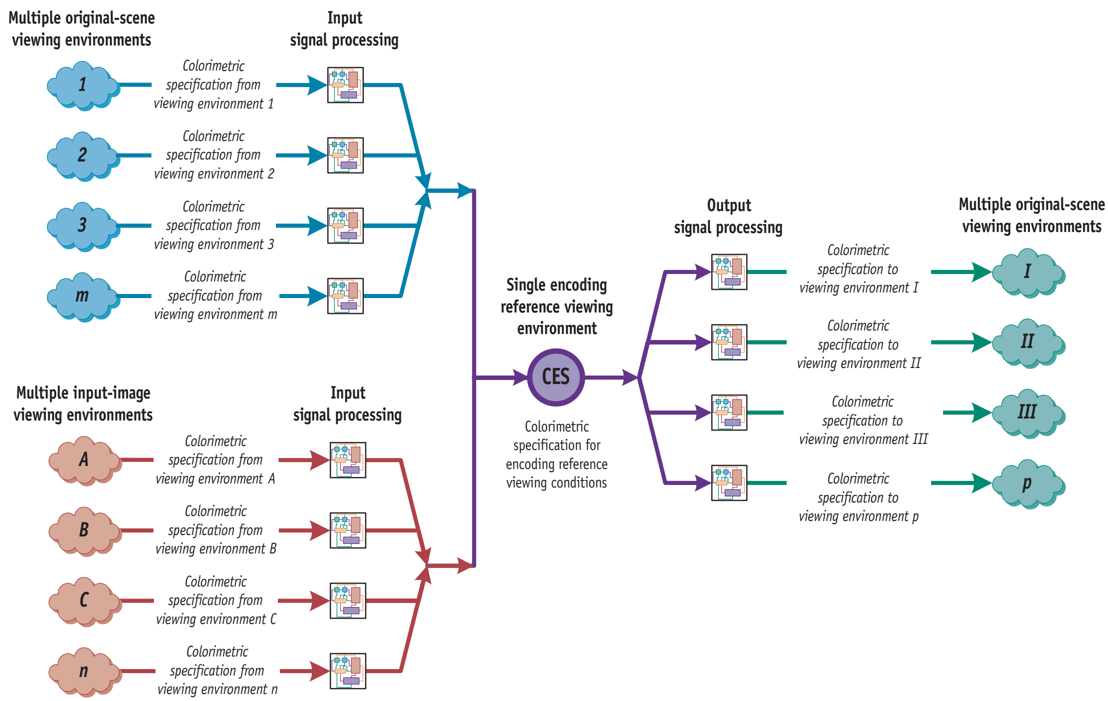


Figure 18.5 The color-encoding strategy of the Unified Paradigm is designed to deal with the many different viewing conditions that can exist for original scenes, input images, and output images.

understand such effects. One result of these efforts has been the development of several sophisticated color-appearance models. These models are extremely useful in many color-related applications, and their development has provided a number of important insights into the nature of human color vision.

A complete discussion and comparison of these complex models is well beyond the scope of this book. Moreover, the paradigm and encoding method described here are not based on any particular color-appearance model; they are based on the *concept* of encoding in terms of color appearance. It is to be expected that the level of understanding of color appearance will continue to rise and that, as a consequence, color-appearance models will continue to become more sophisticated and more accurate. As that happens, those improvements can be incorporated in future implementations of systems based on the Unified Paradigm. For the present, the simple transforms described in the next chapter have been shown to work successfully on a number of experimental and commercial color-imaging systems and should be more than adequate for the vast majority of applications.

Our main goal here is to describe the *objectives* that must be achieved when applying color-appearance models to color encoding; it is not to recommend the use of any particular models or any specific appearance-modeling procedures. However, based on our experience in designing imaging media and systems, and from experience gained in developing a prototype implementation of the Unified Paradigm, we can provide some general guidelines for a straightforward, practical implementation of a color-appearance-based color-encoding method. These guidelines are as follows:

> *For simplicity, both in concept and in implementation, only the most important factors affecting the relationship between colorimetry and color appearance need be considered.*

Although numerous perceptual phenomena have been identified, many of the effects they produce in typical image-viewing situations are small enough to safely ignore. Other effects, such as those induced by adjacent colors, need not be accounted for in most imaging applications. For example, if the presence of one color in an input image affects an observer's perception of an adjacent color, the same phenomenon should occur when an observer views an output reproduction of that image. So there is no real need to account for the perceptual effect in the color encoding. Strictly speaking, then, the encoding describes colors in terms of visually equivalent stimuli viewed in the context of the images in which they reside. In the chapter that follows, the perceptual factors we believe to be most important will be specified, and practical methods of accounting for the effects of each factor will be described.

The magnitudes of color-appearance factors should be consistent with the viewing of images.

Experimental work performed to study visual phenomena sometimes has involved the viewing of color patches and other types of simple color stimuli. Our experience is that quite different results are obtained when more complex stimuli, such as typical pictorial and graphics images, are viewed. In almost all cases, the magnitudes of the effects are reduced for complex stimuli. These reduced magnitudes should be used in the colorimetric transformations.

The magnitudes of all color-appearance factors should be consistent with the color-reproduction properties of actual imaging products.

In particular, it was shown earlier that the chromas of reproduced colors must be higher than those of outdoor-scene colors in order to compensate for the relatively low absolute luminances of reproduced images. However, the chroma compensation required to fully duplicate, indoors, the colorfulness of outdoor-scene colors would be impractically large.

Moreover, in virtually all artwork and in commercial imaging products, the extent of such compensation actually is far smaller. Using a similarly reduced level of compensation for encoding color signals produced directly from live scenes—for signals from digital cameras, for example—results in a consistency between the encoding of such signals and other color signals obtained from direct measurements of reproduced images. This strategy greatly enhances input compatibility, and it produces encoded colorimetric values that realistically can be achieved by actual output devices and media.

Whenever possible, the implementation of the transformations should be mathematically simple, and the transformations themselves must always be reversible.

Mathematical simplicity is a practical issue, especially on systems having limited computational resources. Mathematical reversibility is particularly important in applications where the objective is to make duplicates of input images. In such cases, the input and output signal processing transformations essentially must be mirror images of one another.

Other color-encoding concepts

As discussed in Chapter 17, color-management systems generally are based on just one of the Type *A*, *B*, or *C* paradigms. Such systems are, by definition, subsets of the Unified Paradigm. Accordingly, their underlying color-encoding concepts are less comprehensive than those of the full paradigm. The use of more restricted color-encoding concepts may be perfectly appropriate if the limitations related to their use are fully recognized and judged to be acceptable for a given application. Problems arise, however, when color-management systems based on inherently restricted concepts are promoted as being "universal" solutions for color management—an all-too-common occurrence. It is important, then, to recognize when color-encoding concepts are universal and when they are applicable only under certain conditions.

One principal distinction between the color-encoding concepts of the Unified Paradigm and those of other color-management systems is the difference in what is meant by a "reference viewing environment." In many color-managed systems—those designed for graphic arts applications, for example—the spectral power distribution of a reference viewing illuminant is included in the definition of the CES reference viewing environment. Moreover, a concept central to the associated color management is that the reference viewing conditions apply not only to the encoding but also to the actual viewing of input and output images. Although this approach is consistent with the requirements of particular segments of the imaging industry, imposing a similar limiting interpretation of "reference viewing environment" on the rest of the industry would be contrary to the goals of a universally applicable system. As discussed previously, in most applications users must have the option of determining the illuminants and other parameters of the actual input and output environments, independent of the reference viewing environment defined for the CES.

The concept of "reference viewing environment" is fundamentally different in the Unified Paradigm in that the CES reference viewing conditions and other encoding specifications apply only to the color encoding itself. Unified Paradigm encoding specifications are defined solely to allow the application of appropriate colorimetric transforms needed to retain an intended color appearance into or out of the CES. Consistent with the concepts of a universal system, the encoding specifications do not dictate or restrict what viewing conditions must be used for the actual inputs or outputs of any given system.

Another important distinction between the Unified Paradigm and other color-management approaches is that the color encoding of many systems is based on what is referred to as a "reference rendering medium." In such systems, images are rendered or re-rendered as necessary such that when encoded, they appear to have been produced on a specified reference medium or device. That encoding method can be quite useful, especially in Type *B* paradigm systems where the objective is to make input images look as much alike as possible. It is important, of course, that the colorimetric properties of the reference medium do not overly restrict the encoding capabilities or subsequent functionality of the system. Specifically, the color gamut and luminance dynamic range of the reference medium must be sufficient to encode and subsequently manipulate any anticipated types of images supported by the system. In most cases, that is best accomplished using a hypothetical reference rendering medium. (That technique will be demonstrated in Chapter 21 as part of an example color-management system for digital cinema.) However, in certain applications it may be sufficient for the reference medium to emulate the characteristics of an actual medium. That approach is common in graphic arts systems, for example, where input images typically are re-rendered as necessary to conform to the properties of a representative reflection-print medium.

Although a color-management system based on the colorimetric characteristics of an actual rendering medium can be used successfully for certain applications, problems may arise when imaging media having capabilities greater than those of the CES reference medium are used for input or output. For example, a reference reflection-print medium might be specified as having a maximum reflectance of 90 %, which would correspond to a luminance-factor value of 0.90 (an L^* of about 96), and a CES incorporating that limit would be adequate for encoding images from most types of reflection prints. As discussed in earlier chapters, however, encoding the appearance of other types of images—dark-projected and self-luminous images, for example—often requires the representation of luminance-factor values of 2.00 (an L^* of about 130) or greater. Therefore the actual appearance of these types of images cannot be represented properly in an encoding specification defined in terms of the properties of a realistic reference reflection-print medium. Among other things, this means the CES would not be capable of supporting the input-image-based Type *A* paradigm for certain types of images. Such images could be accommodated in the encoding by re-rendering them to correspond to the properties of the reference medium, but that would impose a Type *B* paradigm outcome. Whether the resulting alteration in image appearance would be acceptable or not would depend on the requirements of the user. The issue here is that on this type of system, no other options may be possible. By definition, then, color-management systems based on any encoding concept that includes a restrictive encoding reference medium cannot be considered universal. In contrast, there is no encoding reference medium associated with the CES of the Unified Paradigm. As a result, the encoding capabilities of its CES are inherently unrestricted.

When considered together, the two issues just discussed illustrate what is perhaps the most basic conceptual difference between many other color-management approaches and the Unified Paradigm: a system based on a fixed approach effectively dictates the rules as to how images will be treated (e.g., that they will be re-rendered, transformed to scene space, etc.), what image-capture and image-viewing environments must be used, and what other rules must be adhered to in order comply with restrictions imposed by the system itself. By comparison, in a Unified Paradigm system decisions regarding issues such as these are made according to the requirements of the user, not according to rules imposed by the system.

Another fundamental conceptual distinction between the color encoding of the Unified Paradigm and that of many other color-management systems is the role of image file tags. We have stated elsewhere in this book that auxiliary tagged information is useful in any color-management system, and we have included such tags in every system we have designed. However, in many other color-management systems, the approach is quite different in that tags are *required* in order for the system to function at all. Very often, tags are employed as a means to address problems resulting from the use of color-encoding concepts that are inherently limited. For example, an input image having L^* values exceeding those of a CES could be identified as such by the inclusion of an image file tag. Then, instead of altering the image by re-rendering it to conform to the CES limits, the original colorimetric information could be preserved by scaling its L^* values differently than those of other types of images. However, because that different scaling would be applied only to some images and not to others, CES image values alone would become ambiguous, and tagged information would be required in order to interpret their meaning.

In this tag-dependent concept of color encoding, multiple distinct and incompatible populations of CES image values are created, and files from the different populations cannot simply be treated as being interchangeable. Unless additional population-specific signal processing is subsequently applied, images from various populations cannot be composited, edited, or manipulated exactly the same way, nor can they be processed through the same output-device transforms. In effect, then, systems based on this approach comprise multiple subsystems. Because they are not inherently united by a single CES, these subsystems operate essentially in parallel rather than as parts of a single integrated entity. Whether a tag-dependent mode of operation is acceptable or not once again depends on the requirements of the system users. Our experience is that for the most part, systems based on this concept ultimately fail. As new circumstances create more and more “exceptions” to the initial rules, more and more tag definitions must be added and supported. Eventually, these systems often reach levels of such complexity that they become unworkable.

Although it should be expected that most systems will evolve over time, the success of that evolution requires initial designs that are inherently extensible and that correctly anticipate the ultimate states of the systems. If the intended ultimate state of a system is one of universal applicability, the underlying concepts of the original design must not be inherently restrictive. In particular, the encoding concept of the CES must be such that new image types, image interpretations, and system functionalities can be added as needed and readily integrated into the system.

Of course the ultimate objective of the Unified Paradigm is, by definition, to be universally applicable. For that reason, the basic color-encoding concept of the paradigm allows all images and image interpretations, regardless of the degree of disparity, to be represented unambiguously and compatibly in a single CES. When that CES is designed according to the color-encoding concepts that have been described here and in previous chapters, a solid foundation is created upon which color-imaging systems that are inclusive, extensible, and truly universal can be constructed.

Discussion

This chapter has described the basic principles of a method for representing color in a way that is consistent with the requirements of individual imaging systems that are based on the Unified Paradigm. The combination of a colorimetric specification and a set of encoding reference viewing conditions makes the color representation unambiguous and universal. Because the representation is of color itself, it is not subject to any restrictions that might be imposed by a particular interpretation, context, medium or device. With the use of appropriate color transformations, as described in the chapter that follows, the appearance of any given color, regardless of its origin or ultimate destination, can be represented.

Lastly, it is important to emphasize that although this color-encoding method solves the basic problem of *representing* color in a unified way, it alone does not create the various types of color *interpretations* supported by the Unified Paradigm. In order to provide multiple input interpretation and output interpretation options, other color-encoding methods also must be incorporated as part of the image signal processing. For example, a scene-based color-encoding method might first be used to transform color values measured from an input device or medium to scene-space colorimetry. Appropriate colorimetric transforms then would be used to transform that colorimetry to color-appearance values for encoding in a CES. Additional color-encoding methods, such as those based on re-rendering to reference media or display devices, also must be included in the overall design in order to generate other constituent-paradigm outcomes.

An appearance-based color-encoding method, then, is not a *replacement* for the other color-encoding methods discussed in Part III. Instead, it is a means for achieving input compatibility in a way that can support the incorporation of multiple color-encoding methods. Examples of the use of alternative color-encoding methods in combination with a color-appearance-based CES will be shown in Chapters 20 and 21.

Summary of key issues

- In the Unified Paradigm, each alternative interpretation of color from each input device or medium constitutes a unique input source.
- The appearance-based color-encoding method of the Unified Paradigm supports all input- and output-image sources.
- Unambiguous, color-appearance-based encoding can be achieved by the use of colorimetric values together with a defined set of encoding reference viewing conditions.
- In the Unified Paradigm, the specifications for the encoding reference viewing conditions apply only to the color encoding itself. The conditions for original-scene, input-image, and output-image viewing are entirely independent of the encoding reference conditions.
- The underlying philosophy and encoding concepts of most color-management systems are fundamentally different from those of the Unified Paradigm. These differences generally limit the functionality of such systems to those of just one of the constituent paradigms of the Unified Paradigm.
- Color encoding based on color appearance does not, in and of itself, provide all of the features of the Unified Paradigm. Other encoding methods must be used in combination with appearance-based encoding in order to provide alternative interpretations of input or output images.

19

A Unified Paradigm: Encoding Transformations

The preceding chapter introduced the basic principles of an appearance-based color-encoding method that is consistent with the objectives of the unified color-management paradigm. In this chapter, the colorimetric transformations required for implementation of that encoding method will be described. As was discussed previously, these colorimetric transformations are necessary to account for any differences of the original-scene, input-image, and output-image viewing conditions from the encoding reference viewing conditions.

It is important to emphasize that the colorimetric transformations do *not* determine how the *appearance* of a color stimulus, specified in terms of its colorimetry, would *change* if it were moved from one set of viewing conditions to another. The transformations do just the opposite; they determine how the colorimetry of the stimulus *itself* must be modified such that its color appearance is *maintained* in each different set of viewing conditions.

Although other factors also might be considered, our experience is that excellent appearance-based color encoding can be achieved if the colorimetric transformations account for each of the following factors:

- Viewing-flare contribution to observed stimuli
- Stimuli absolute luminances
- Observer chromatic adaptation
- Observer lateral-brightness adaptation
- Observer general-brightness adaptation.

In order to appropriately account for these factors, each set of image-capture, input-image, and output-image viewing conditions must be fully characterized in terms of viewing flare, observer state of chromatic adaptation, image absolute luminances, and image surround type. Appropriate colorimetric transformations then must be developed based on the relationships of those conditions to the encoding reference conditions. The components of the viewing-condition characterizations, and some example methods that can be used for determining the colorimetric transformations, will now be described.

Viewing-flare transformations

The presence of viewing flare in an environment will physically alter the color stimuli—illuminated objects of live scenes, illuminated hardcopy images, self-luminous softcopy images, or projected images—viewed in that environment. In order to properly account for such alterations, the amount and chromaticity of the flare light in each set of

Digital Color Management: Encoding Solutions, Second Edition Edward J. Giorgianni and Thomas E. Madden
 ISBN: 978-0-470-51244-9

viewing conditions must be defined. The amount of flare usually is specified relative to the amount of light from a stimulus that would be perceived as white in the viewing conditions.

If viewing flare is different in two sets of viewing conditions, the effects of that difference must be compensated for by an appropriate colorimetric transformation. For example, if it is desired to encode the appearance of an input image as viewed in conditions in which the amount of viewing flare is greater than that specified for the encoding reference conditions, the colorimetric effects of that additional flare light must be included in the input colorimetric transformation. The effects of increased flare light on a system grayscale and on a small set of test colors are shown in Figures 19.1a and 19.1b, respectively.

The colorimetric results of viewing flare shown in the figures can be computed using the principles of additive color mixing. The first step in this process is to determine the tristimulus values (X_{fl}, Y_{fl}, Z_{fl}) for the flare light, based on the specification of that light in terms of its chromaticity coordinates (x_{fl}, y_{fl}, z_{fl}) and its percentage (pf) of the image white luminance (Y_w). The tristimulus values of the flare light can be computed as follows:

$$
\begin{aligned}
Y_{fl} &= \frac{pf}{100} Y_w \\
X_{fl} &= \frac{x_{fl}}{y_{fl}} Y_{fl} \\
Z_{fl} &= \frac{z_{fl}}{y_{fl}} Y_{fl}
\end{aligned}
\qquad (19.1a)
$$

For each color stimulus, a modified stimulus is computed by adding the tristimulus values for the flare light to the tristimulus values for the color stimulus, as shown below:

$$
\begin{aligned}
X_f &= X_{nf} + X_{fl} \\
Y_f &= Y_{nf} + Y_{fl} \\
Z_f &= Z_{nf} + Z_{fl}
\end{aligned}
\qquad (19.1b)
$$

where X_{nf}, Y_{nf}, and Z_{nf} are the tristimulus values for a stimulus with no flare light, and X_f, Y_f, and Z_f are the tristimulus values for a stimulus with flare light added. These and other viewing-flare calculations are described in greater detail in Appendix E.

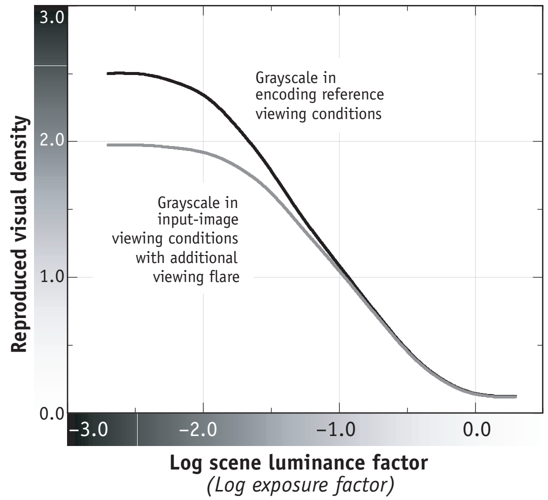


Figure 19.1a The colorimetric effects of increased viewing flare on a system grayscale.

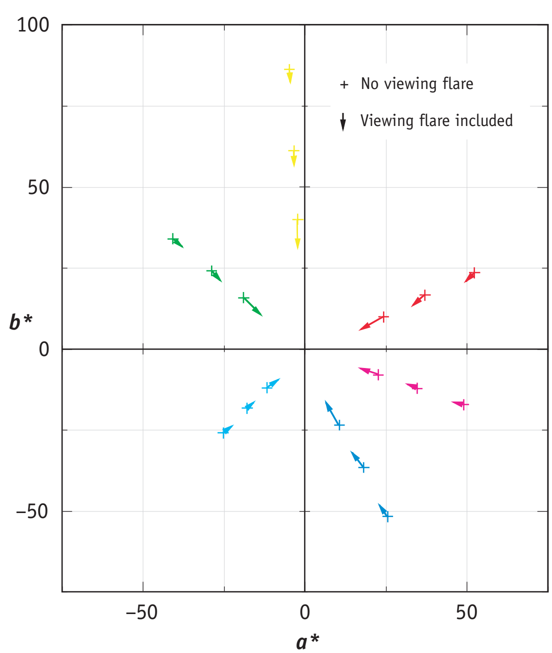


Figure 19.1b The colorimetric effects of increased viewing flare on a set of test colors.

Luminance-level transformations

As was discussed previously, the absolute luminances of stimuli will affect an observer's perception of luminance contrast and colorfulness. In order to account for such effects, the luminance levels of the stimuli being viewed must be defined. This can be done by specifying, for each set of viewing conditions, the luminance of a stimulus that would be perceived as white.

While the luminance levels of typical viewed input and output images certainly may differ, the effects of those differences often are small enough to ignore. However, the luminance levels of outdoor scenes generally are very different from those of most reproductions viewed indoors. The effects of those large differences *must* be accounted for in the colorimetric transformations.

One procedure for determining a luminance-level transformation is described in detail in Appendix G. The transformation relates scene-space colorimetric values and reproduced colorimetric values in a way that is realistic and consistent with the color-reproduction properties of practical imaging media. Figure 19.2a shows the effect of the colorimetric transformation on a system grayscale characteristic, and Figure 19.2b shows the effect on a set of test colors.

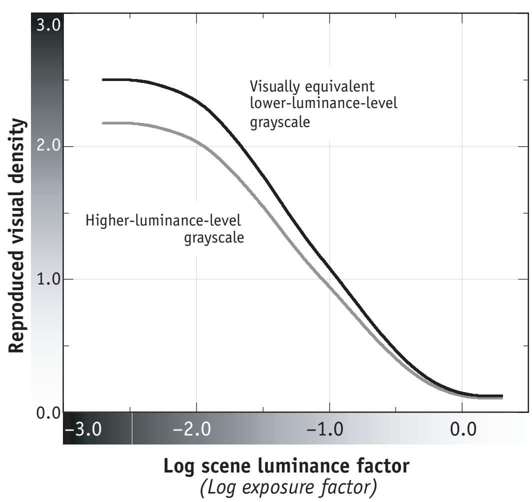


Figure 19.2a The effect of a luminance-level transformation on a system grayscale characteristic. The transformation accounts for the fact that, for visual equivalence, higher grayscale–luminance contrast is required at lower luminance levels.

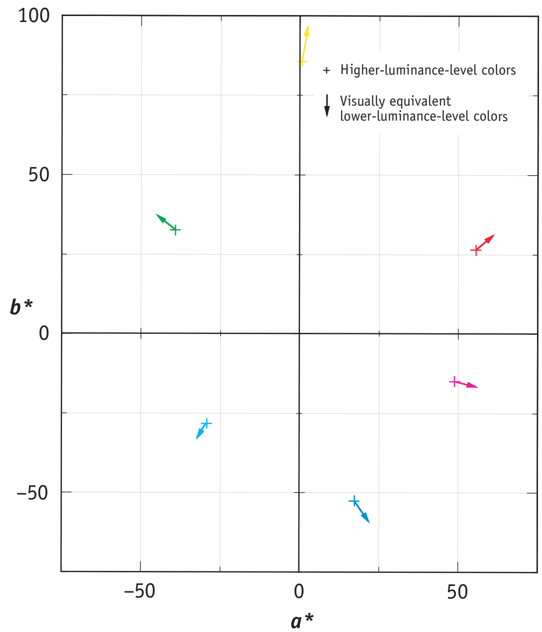


Figure 19.2b The effect of a luminance-level transformation on a set of test colors. The transformation accounts for the fact that, for visual equivalence, higher chromas are required at lower luminance levels.

Chromatic adaptation transformations

The perception of color stimuli is strongly affected by the observer's state of chromatic adaptation. That adaptive state therefore must be specified for each set of viewing conditions. The specification can be expressed in terms of an *observer adaptive white*.

An observer adaptive white defines the chromaticity of a color stimulus that an observer, who is adapted to the viewing conditions, would judge to be perfectly achromatic and to have a luminance factor of unity. It is very important to emphasize that it is only necessary to define an adaptive white in terms of its chromaticities; it is not necessary (nor is it useful) to define an adaptive white stimulus in terms of its spectral power distribution.

If the chromaticities of the adaptive whites are different in two sets of viewing conditions, the perceptual effects due to that difference must be accounted for by the use of an appropriate chromatic adaptation transformation. The transformation determines corresponding tristimulus values of a color stimulus that would produce, for a standard observer chromatically adapted to one set of viewing conditions, a visual match to another color stimulus viewed by a standard observer who is chromatically adapted to a different set of viewing conditions.

For example, Figure 19.3 compares the colorimetric values for a set of stimuli that would produce, for an observer chromatically adapted to the chromaticity of D_{50}, a visual match to another set of stimuli viewed by an observer who is chromatically adapted to the chromaticity of D_{65}. The chromaticities for the visually equivalent stimuli of the figure were determined using the simple von Kries transformation below:

$$\begin{bmatrix} X_{D_{50}} \\ Y_{D_{50}} \\ Z_{D_{50}} \end{bmatrix} = \begin{bmatrix} 1.0161 & 0.0553 & -0.0522 \\ 0.0060 & 0.9956 & -0.0012 \\ 0.0000 & 0.0000 & 0.7576 \end{bmatrix} \begin{bmatrix} X_{D_{65}} \\ Y_{D_{65}} \\ Z_{D_{65}} \end{bmatrix} \tag{19.2}$$

More advanced methods, which account for incomplete adaptation and/or for nonlinearities in the adaptation process, also can be used. However, the methodology of all such methods basically is similar:

1. CIE *XYZ* tristimulus values for a color stimulus first are transformed into three retinal cone-response tristimulus values.
2. The cone-response tristimulus values then are adjusted to account for observer chromatic adaptation effects.
3. The adjusted cone-response tristimulus values are transformed back to (modified) CIE *XYZ* tristimulus values.

In a von Kries transformation, the combined effect of these individual transformations is performed by a single matrix. A procedure for developing a von Kries transformation matrix is described in Appendix D.

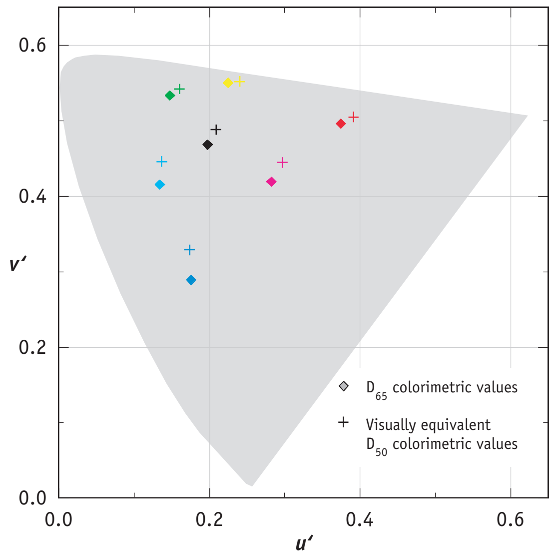


Figure 19.3 Comparison of the colorimetric values for a set of stimuli that would produce, for an observer chromatically adapted to the chromaticity of D_{50}, a visual match to the corresponding set of stimuli viewed by an observer who is chromatically adapted to the chromaticity of D_{65}.

Lateral-brightness adaptation transformations

In certain viewing situations, an observer's perception of color stimuli will be influenced by lateral-brightness adaptation effects induced by the surround in which the stimuli are viewed. To account for such effects, each set of viewing conditions must be characterized in terms of *surround type*. The specification for surround type defines the characteristics of the visual field immediately surrounding the viewed stimuli. The surround can be specified in terms of its chromaticity and luminance factor.

If the chromaticity of the surround is similar to that of the adaptive white, but the luminance factor of the surround differs from the average luminance factor of the viewed stimuli, the surround principally will influence an observer's perception of luminance contrast and overall brightness. This influence of the surround on perceived luminance contrast can be accounted for by using the procedure described in Appendix D.

As shown below, there are two basic steps to that procedure. First, the Y tristimulus value for each stimulus is modified according to an experimentally determined power factor, S, to form a visually equivalent tristimulus value, Y_s, for the different surround:

$$Y_s = Y^S \tag{19.3a}$$

The X and Z tristimulus values for the stimulus then are scaled by the ratio of the Y_s tristimulus value to the unmodified tristimulus value, Y:

$$\begin{aligned} X_s &= X\frac{Y_s}{Y} \\ Z_s &= Z\frac{Y_s}{Y} \end{aligned} \tag{19.3b}$$

This scaling produced by Equation (19.3b) maintains the chromaticity values (x, y) for the stimulus. In certain other lateral-brightness adaptation transformation procedures, the chromaticity values are adjusted such that the chroma of the stimulus also is modified. However, our experience is that for pictorial images, such chroma adjustments are not appropriate.

Applying a lateral-brightness adaptation transformation from darker-surround viewing conditions to lighter-surround viewing conditions results in a decrease in overall grayscale slope (photographic gamma), as shown in Figure 19.4. The consequence of this transformation is that an image having a grayscale of lower slope would be required in lighter-surround viewing conditions in order to visually match the luminance contrast of an image viewed in darker-surround conditions. Conversely, a lateral-brightness adaptation transformation from lighter-surround conditions to darker-surround conditions produces the required increase in overall slope of the system grayscale.

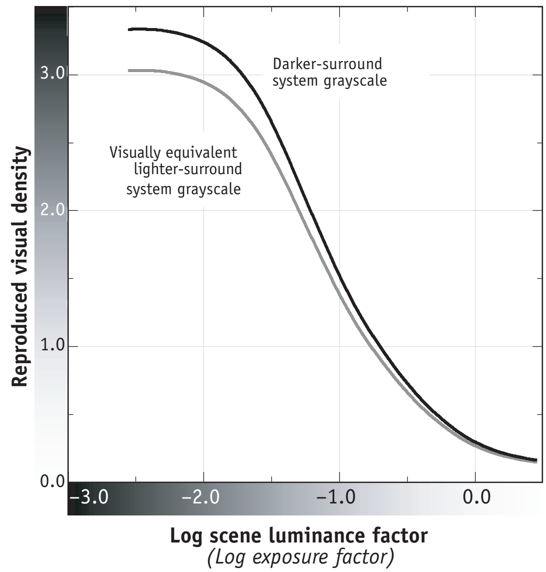


Figure 19.4 Results of applying a lateral-brightness adaptation, or "dark-surround," transformation to a grayscale characteristic. The transformation accounts for the fact that an image having a grayscale of lower slope would be required in a lighter-surround condition in order to match the luminance contrast of an image viewed in a darker-surround condition.

General-brightness adaptation transformations

The previous transformation accounted for the influence of surround conditions on perceived luminance contrast. Surround conditions also may influence the *brightness* of displayed stimuli. This effect, a form of general-brightness adaptation, is most likely to occur in darker-surround viewing conditions. The observer tends to adapt to the overall luminance level of the displayed stimuli themselves because there are few visual cues to otherwise influence brightness adaptation.

Figure 19.5 shows the measured visual densities—which are equivalent to *standard* colorimetric values based on a perfect white reference—for four media:

(a) A conventional photographic reflection-print medium, which has a minimum visual density (or negative log luminance-factor value) of about 0.10.

(b) A particular type of instant-print film, which produces reflection images having a minimum reflection visual density of about 0.20.

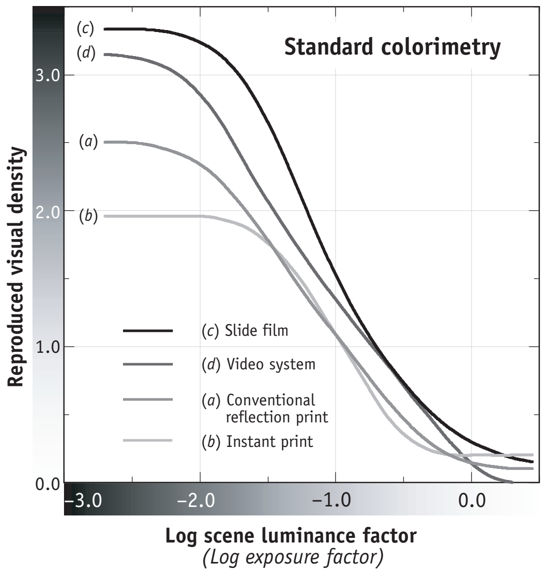


Figure 19.5 Grayscales for four media, measured in terms of standard densitometry (visual density) or standard colorimetry (negative log luminance factor) as a function of scene log exposure factor or scene log luminance factor.

(c) A conventional photographic slide film, which has a minimum transmission visual density of about 0.15.

(d) A high-quality video display, such as a home theater display, which has a minimum negative log luminance-factor value of 0.00 when the measuring instrument is referenced to the stimulus produced from maximum video RGB control signals.

Appearance color encoding of the media shown in the figure, and all other types of media, requires proper accounting for the observer's general-brightness adaptation. One simple procedure for doing that is shown below. First, the CIE Y tristimulus value for each stimulus is adjusted by an experimentally determined scale factor, B, to form a brightness-adjusted tristimulus value, Y_b:

$$Y_b = BY \tag{19.4a}$$

The X and Z tristimulus values for each stimulus then are scaled by the ratio of the Y_b tristimulus value to the unadjusted tristimulus value, Y:

$$\begin{aligned} X_b &= X\frac{Y_b}{Y} \\ Z_b &= Z\frac{Y_b}{Y} \end{aligned} \tag{19.4b}$$

The effect of this general-brightness adaptation transformation is an overall visual density shift of the grayscale. The photographic gamma (slope) of the curve is not affected. Because the same factor is applied to the X, Y, and Z tristimulus values, the chromaticity values for the stimulus are maintained.

In Figure 19.6, the grayscales for the four media of Figure 19.5 are shown in terms of visual density values that have been adjusted for general-brightness adaptation according to the procedure just described. The magnitudes of the brightness adjustments (values for the scale factor B) were determined from visual judging experiments. In these

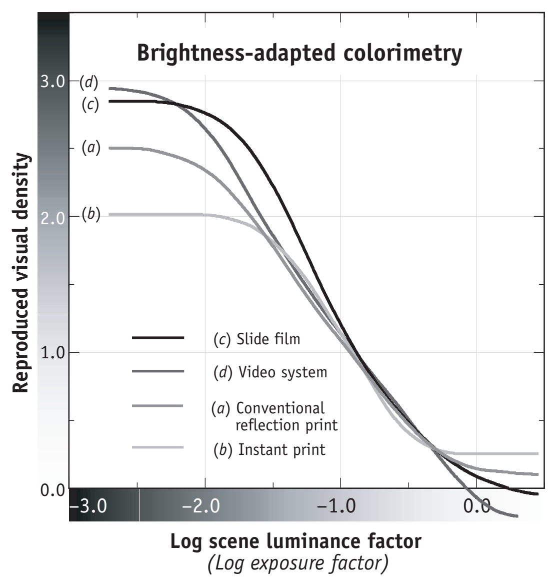


Figure 19.6 Grayscales for four media, expressed in terms of brightness-adapted densitometry or colorimetry. Output images made according to these values will have a correct brightness relationship.

experiments, images from various input media were encoded, the encoded values were output to a single device or medium, and the resulting output images were viewed simultaneously. The encoding included colorimetric transformations for viewing flare, chromatic adaptation, and lateral-brightness adaptation. The value of the brightness-adaptation scale factor was varied in the encoding of images from each input medium until the output images from all input media matched in brightness as closely as possible. The scale factor value for the two reflection media was 1.00; the values for the slide film and the video display were 1.53 and 1.61, respectively.

Due to the different luminance dynamic ranges and other color-reproduction differences of the various input media, the output images never were made *identical* by these brightness adjustments. Nor should they have been. Although that result could have been obtained by *re-rendering* each input to achieve a single color-reproduction outcome, that was *not* the intent. The intent was to produce a set of images that matched in overall brightness but also matched the different color appearances of actual images on the various input media.

As an analogy, consider a comparison of two audio loudspeakers—a small bookshelf speaker and large multi-speaker system. The two will not sound *identical* because of the greater frequency range that can be reproduced by the larger speaker system. However, they can be made to sound equally *loud* by making proper adjustments of the amplifiers. A point will be reached where no further improvements in the match can be made by loudness adjustments alone, and it is at that point that the speakers can be compared most fairly.

When output images from different input media are encoded in terms of their own color appearances and adjusted for overall brightness, a similar point is reached when the matching cannot be improved further by brightness adjustments alone. That point can be used to determine the scale factor for calculating brightness-adapted colorimetric values.

Note that two of the brightness-adapted grayscales in Figure 19.6 have some visual densities *less than* zero. These negative visual densities correspond to luminance-factor values greater than 1.0, and to CIE Y tristimulus values and CIELAB L^* values greater than 100. That is as it should be. When high-quality images are viewed under appropriate conditions, they will *appear* to have areas that are brighter than a reference white. The encoding of such image areas in terms of color appearance *should* have luminance-factor values greater than that of a perfect white reference. This is an important outcome that must be considered in the development of data metrics associated with appearance-based color encoding specifications. For example, in order to appropriately represent the media of Figures 19.5 and 19.6, the visual density range encodable by the data metric must extend to about -0.3, which corresponds to a luminance-factor value of about 2.00 and a CIELAB L^* value of about 130.

Encoding transformation comparisons

As discussed in Chapter 18, the color-encoding concepts of most color-management systems—including systems described as having CESs based on color-appearance encoding—generally differ significantly from those of the Unified Paradigm. Such differences may not be immediately apparent because in the descriptions of their color-encoding transformations, other systems often use terminology similar to that of the Unified Paradigm. However, as discussed below, a comparison of the color-encoding transformations commonly used on other systems to those of the Unified Paradigm reveals a number of fundamental conceptual differences in encoding intents and system design philosophies.

Viewing flare

The viewing-flare transformations used in most appearance-based CESs are generally quite similar to those of the Unified Paradigm. In most cases, colorimetry is specified in terms of flareless measurements. Additional specifications of the relative amount of flare light present in the input, output, and encoding reference viewing environments are included so that the colorimetric effects due to flare light can be computed.

Absolute luminance

Most systems require a specification of the absolute luminance level for input and output images, and most specify a reference luminance level for encoding. However, in most cases, no specific recommendations are made for dealing with input or output luminance levels that differ from those of the encoding environment. Some systems do make mention of the use of color-appearance models. However, our experience is that when large differences in luminance are involved, such models yield results that are both impractical and undesirable for pictorial images. In particular, chroma modifications determined for transformations between typical outdoor and indoor luminance levels are entirely unrealistic for pictorial imaging applications. That result should be expected because the objectives are quite different. Color-appearance models are designed to predict color *matching*. The Unified Paradigm system transformation procedures described earlier in this chapter are designed to produce preferred color *renditions*; thus they are more appropriate for the intended purpose.

Chromatic adaptation

Although virtually all color-management systems address the issue of chromatic adaptation, in most cases the approaches used are fundamentally different from that of the Unified Paradigm. These differences provide a clear indication of how dissimilar various systems can be in their underlying philosophies.

Earlier in this chapter, we defined the term "observer adaptive white" for the Unified Paradigm as "the chromaticity of a color stimulus that an observer, who is adapted to the viewing conditions, would judge to be perfectly achromatic and to have a luminance factor of unity." It is important to recognize that this definition refers to the observer's *reaction* to an environment; it does *not* define a property of the environment itself.

For example, in a particular viewing environment an observer may be adapted such that stimuli having the chromaticity of CIE Standard Illuminant D_{50} appear to be achromatic. However, there are many factors that affect an observer's state of chromatic adaptation, including mixed lighting from multiple illuminants and incomplete adaptation to the environment. Consequently, the actual source of illumination in this example may or may not correspond to the spectral power distribution, or even to the chromaticity, of CIE D_{50}.

In the CES of the Unified Paradigm, a clear and explicit distinction is made between an image illuminant and an observer's state of chromatic adaptation. The contribution of an actual image illuminant is understood to be *already* incorporated in the colorimetric values of the image data. This distinction is particularly relevant for self-luminous images, where there is no separate viewing illuminant; the color stimuli are formed solely by light produced by the display color primaries. In the Unified Paradigm, then, observer adaptive white chromaticities are used only to allow the computation of colorimetric transforms necessary to *maintain* a stated color appearance, as described by the image colorimetry and its particular associated viewing environment.

In many other systems, however, the explicitly stated or apparently assumed interpretation of a reference viewing environment is entirely different: it prescribes a set of conditions *under which the image is to be viewed*. This means that the objective of the chromatic adaptation transformation is not to determine the colorimetry necessary to *maintain* a stated color appearance. Instead, the intent (whether achievable or not) is to *change* the colorimetry to what would have been formed had the image illuminant been that of the reference environment. (The distinction between these two objectives was discussed at some length in Chapters 15 and 18.) This interpretation of course requires a specification of the *complete spectral power distribution* of the reference illuminant, in addition to its chromaticity. A CES based on a reference illuminant defined in terms of a spectral power distribution can be valid for certain types of systems. However, it clearly cannot serve the function of a universal color specification that must support various types of input and output viewing environments and various forms of images. For example, there would be a fundamental ambiguity as to how self-luminous images, used either for input or for output, could be related to a reference-illuminant-based CES.

The importance of the philosophical difference between the treatment of chromatic adaptation in the CES of the Unified Paradigm and that of other system cannot be overstated. It once again clearly illustrates the fundamental distinction previously drawn in Chapter 18: in the Unified Paradigm, the *user* determines how images will be interpreted throughout the imaging chain; in systems that specify that image colorimetry is to be based on reference illuminants, image interpretation becomes dictated by the system itself.

Lateral-brightness adaptation

In many color-management systems and color encoding specifications, image-surround conditions are not explicitly defined. Quite often, however, an average surround—described, for example, as "an achromatic neutral of 20 % reflectance"—is assumed. This and similar statements are consistent with the reflection-print-based paradigm underlying many color-management systems. However, the statement once again raises questions regarding the applicability of these systems to other forms of images. We recall one person who was attempting to use a particular color-management system asking, "Since I'm dealing with images on a monitor, what exactly am I supposed to do with the 20 % neutral reflector?"

While reflection-print-based color-management systems generally include no specific recommendation for dealing with non-reflective images or non-reflective image surrounds, both are fully supported in the Unified Paradigm system. Throughout the system and in its CES, image surrounds are defined in universally applicable terms specifying their *luminance* and *chromaticity* relative to the observer adaptive white for the associated input, encoding, or output viewing environment. This allows images of *any* form having *any* type of surround to be properly represented in terms of their appearance at any stage of the imaging system.

Some other systems do recognize the need to support images associated with different types of surrounds, and they include some form of colorimetric transformation for use when the surround of an actual input or output image differs from that of the encoding specification. Most often, these transformations are based on color-appearance models. However, our experience is that for pictorial images, most such models yield poor results. They tend to overcorrect luminance contrast, and they also produce inappropriate adjustments in chromas. The simple procedures discussed earlier in this chapter generate far more desirable results.

General-brightness adaptation

In the CES of the Unified Paradigm, the transformation method described previously in this chapter is used to produce *brightness-adapted* colorimetric values in the CES. As was discussed, this method results in images that are equivalent in perceived overall brightness when shown in sequence or when displayed simultaneously on any given output device or medium. Consistent with the principles of the Unified Paradigm, this method retains the specified appearance of images; it does not force all images (e.g., by re-rendering) to have the same basic appearance.

It is often suggested that general-brightness adaptation is automatically accounted for in other systems by the use of what is sometimes referred to as "relative colorimetry" or, more appropriately, "media-relative colorimetry." However, we strongly disagree. The distinction between the use of media-relative colorimetry and the use of general-brightness adaptation transformations in the CES of the Unified Paradigm is extremely important and warrants the extensive discussion that follows.

Many colorimetric calculations yield values that are expressed relative to those of a reference white. For example, in terms of CIELAB values, the reference white always will have an L^* value of 100.0 with a^* and b^* values of 0.0, regardless of the reference's absolute luminance or chromaticity. However, in CIE standard colorimetric calculations, the reference white is not arbitrary: it is defined explicitly as a *perfect white* (i.e., an ideal isotropic diffuser with a spectral reflectance factor or spectral transmittance factor equal to unity at each wavelength of interest). In "media-relative" colorimetry, however, something other than a perfect white—usually the support of the particular medium being measured—is used as the reference white.

There are situations where the use of such colorimetry may be appropriate. In certain graphic arts applications, for example, media-relative colorimetry essentially provides a description of the colorimetry of the ink image, independent of the paper on which it is printed. That colorimetry is useful when the intent is to produce the same ink image (rather than the same color appearance) on a number of different papers. The use of media-relative colorimetry can also help minimize certain undesirable printing artifacts, generally associated with lighter colors, that are unique to halftone printing.

It is our position that because they are specific to particular output technologies, halftone printing problems and other such issues should not be a principal factor in determining the basic color-encoding methodology for a system intended to support an entire imaging industry. Instead, the appropriate place to address output-specific problems is in the output signal processing transforms associated with the relevant devices.

In the Unified Paradigm, the option to use media-relative colorimetry is handled exactly that way—as an option that can be incorporated in the output signal processing transforms of particular devices for which the technique might be useful. Unlike many color-management systems, the Unified Paradigm does *not* use media-relative colorimetry in the CES that is the very heart of the system. The reasons for this can be demonstrated by a series of examples using the four media described previously in this chapter.

In the computation of media-relative colorimetry of reflection images, the colorimetry of the paper support generally is used as the white reference. The underlying assumption is that observers adapt to the support because it is the brightest part of an image on that medium. In reality, however, reflection images are judged essentially as *objects* within the viewing environment. In typical reflection-image environments, there is little or no adaptation to a reflection image because the viewing environment itself contains white objects or other visual cues that strongly influence the adaptive state of the observer.

Our experience is that when the objective is to encode reflection images in terms of their color appearance, standard colorimetric values based on a *perfect* white reference should be used. In Figure 19.7, for example, the grayscales for the photographic and instant-print media of Figure 19.5, measured relative to a perfect white diffuse reflector, show that the minimum visual densities of the media are different. That is consistent with how images on these media actually appear. In particular, the higher minimum visual density of the instant print is plainly visible under almost any reasonable set of viewing conditions. If standard colorimetric values of images on these two input media are encoded and output to the same device, the output images will have a correct brightness relationship. So if the output device is set up such that it produces images of proper brightness from one of the input media, the device also will produce images of proper brightness from the other input medium.

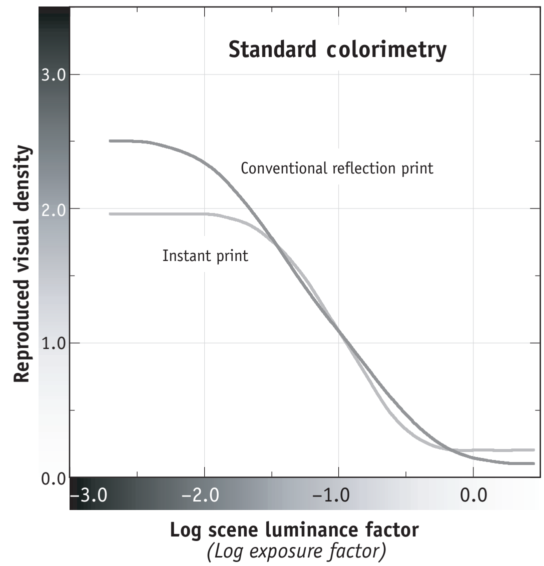


Figure 19.7 Grayscales for two reflection media, measured in terms of standard densitometry or standard colorimetry. Output images made according to these values will have a correct brightness relationship.

In Figure 19.8, the *media-relative* grayscales for these two media are shown. Since media-relative values are based on respective minimum values, each grayscale has a minimum media-relative reflection density value of *zero*. These grayscale values therefore imply that the minimum density areas of the media are visually identical to each other and that they are visually indistinguishable from a perfect white reflector. But that is *not* how they actually appear. In reality, the differences in their minimum density

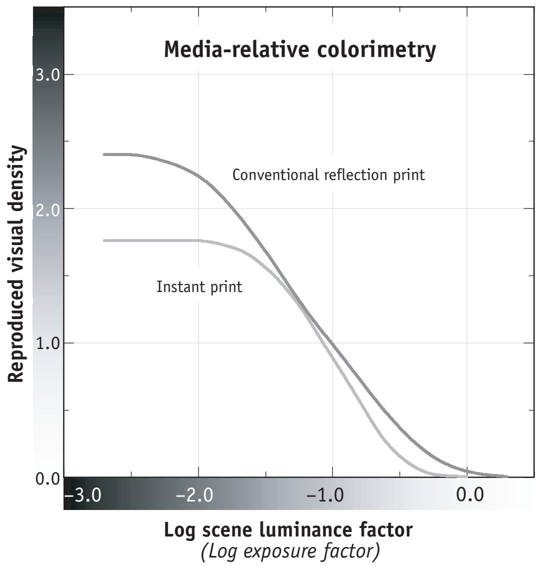


Figure 19.8 Grayscales for two reflection media, measured in terms of media-relative colorimetry. Output images made according to these values will *not* have a correct brightness relationship.

areas are very apparent, and neither medium has areas that appear to be perfectly white.

Moreover, the use of media-relative computations has reduced *all* the density values of the instant-print medium to a much greater extent than it has reduced those of the photographic medium. If the media-relative values from both inputs were encoded and output to the same device, the resulting images would *not* properly represent the brightness relationships of the input images. Instead, output images from instant-print input would be too light. If the output were adjusted to correct that problem, output images from conventional photographic print input then would be too dark. In the case of reflection images, then, the use of media-relative colorimetry only serves to *create* a problem in an attempt to "fix" a problem with standard colorimetry that, in reality, never existed.

The situation is somewhat different when media-relative colorimetry is used for color encoding other types of media. For example, in the computation of the media-relative colorimetry of transparency images, such as photographic slides or motion picture projection prints, the minimum transmission visual density of the medium is used as the white reference. Again, the underlying assumption of relative colorimetry is that observers adapt to the brightest part of an image. As was discussed earlier, there *is* a significant amount of adaptation to projected images, especially when they are viewed in completely darkened environments. However, what is perceived as a perfect white seldom if ever corresponds to the *brightest* areas of these images. As discussed in Chapter 7, the grayscale characteristics of transparency media are designed such that the brightest areas appear to be *brighter* than a perfect white.

On the other hand, because some general-brightness adaptation does occur, the encoding of projected images cannot simply be based on a standard white reference. Doing so results in standard density values and standard colorimetric values, but those values do *not* represent the perceived brightnesses of images. Figure 19.9, for example, shows the grayscales for the photographic reflection print and slide media of Figure 19.5, as measured by standard

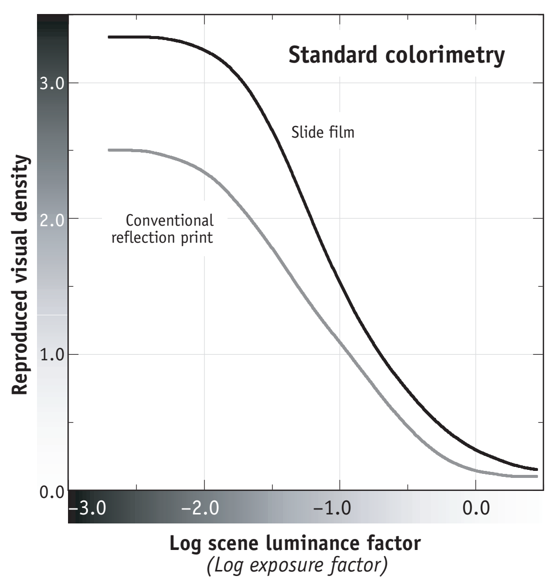


Figure 19.9 Grayscales for a photographic reflection print and slide film measured in terms of standard densitometry or colorimetry. Output images made according to these values will *not* have a correct brightness relationship. Images made from slide-film input will be darker than those made from reflection-print input.

densitometric and colorimetric methods. The figure shows that, at any given exposure-factor value, the visual densities of the slide film are higher than those of the reflection print. This implies that projected slide images must be darker overall than reflection-print images. But, of course, that is not how images on these media actually compare. Slide films have been manufactured and sold for more than 70 years. If it were true that they always produced images that were too dark, someone certainly would have noticed by now!

In this particular example, the use of media-relative colorimetry (Figure 19.10) will not greatly change the relationship described by standard colorimetry because the minimum visual densities of the two media only differ by 0.05 density units. If values based on *either* standard *or* media-relative measurements were encoded and output to the same device, the resulting images would not have a correct brightness relationship. Images produced from slide-film input would be much darker than those from reflection-print input. In this example, then, the use of media-relative colorimetry was intended to fix a problem that indeed existed, but its use actually had very little effect.

Figure 19.11 shows the grayscales for another pair of media, the photographic reflection print and the high-quality video display, from Figure 19.5. If images are made from this standard colorimetry, the images made from video display values will be darker than those made from reflection-print values.

Figure 19.12 shows the media-relative grayscales for these media. These grayscales now imply that video images are even *darker* compared to reflection-print images than was suggested by the standard colorimetric measurements. As a result, if an output device is set up such that high-quality images are produced from reflection-print input, images produced from media-relative video display values will be much too dark. In this example, then, the use of media-relative colorimetry not only failed to correct a real problem associated with the use of standard colorimetry, but also made that problem considerably *worse*.

Media-relative colorimetry has failed in each of these situations because it does not correlate well

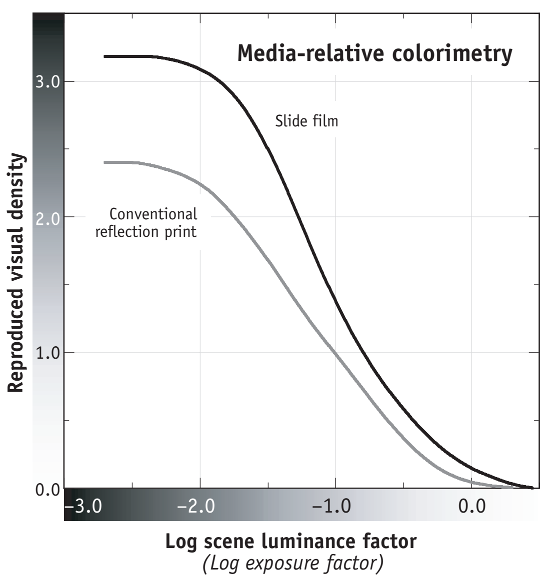


Figure 19.10 Grayscales for a photographic reflection print and slide film measured in terms of media-relative colorimetry. Output images made according to these values again will *not* have a correct brightness relationship.

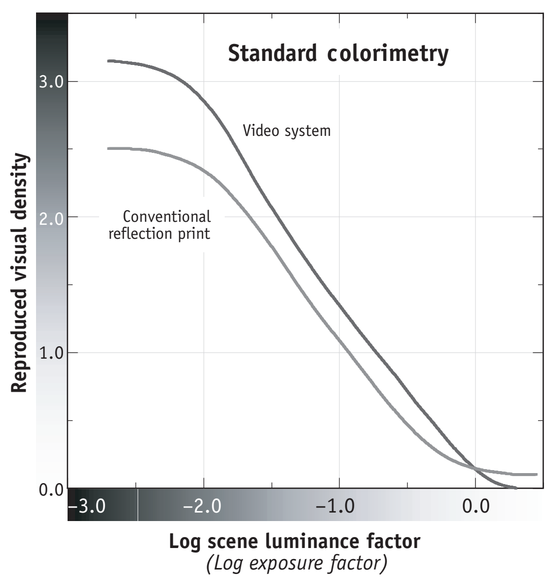


Figure 19.11 Grayscales for a photographic reflection print and a high-quality video display measured in terms of standard densitometry or colorimetry. Output images made according to these values will *not* have a correct brightness relationship.

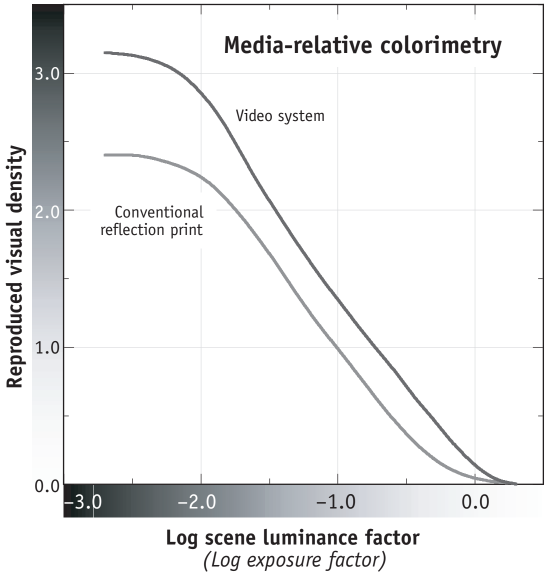


Figure 19.12 Grayscales for a photographic reflection print and a high-quality video display measured in terms of media-relative colorimetric methods. Output images made according to these values again will *not* have a correct brightness relationship.

with actual color appearance. Consequently, encoded color values based on that colorimetry do not correctly describe the perceived relative brightnesses of most image types, especially those having different surround conditions. The meaning of such values therefore is *ambiguous* because the brightness relationships of different types of images cannot be determined from their encoded values alone. By comparison, the use of appropriate brightness-adaptation transformations results in an unambiguous representation of the correct brightness relationships of images from all input sources, as was shown in the examples of Figure 19.6.

Other color-encoding strategies

The manner in which general-brightness adaptation and other colorimetric transformations are interpreted and implemented in most other systems exemplifies the fundamental differences of such systems from the Unified Paradigm. In particular, the most common strategies for achieving interoperability are based on encoding all images, regardless of their actual origins, to look as much alike as possible. That outcome can be imposed, for example, by a required rendering or re-rendering of images to conform to the properties of a specified reference medium viewed in a particular viewing environment. Such systems function according to the Type *B* paradigm rather than as an embodiment of the Unified Paradigm. Again, a color-encoding method that fully supports the Unified Paradigm must allow included systems to function according to *any* of the Type *A*, Type *B*, and Type *C* paradigms, as determined by the wishes of the users.

As discussed in Chapter 18, a common strategy often used to expand the capabilities of a color-management system is to append various data identification tags to the image files. This strategy is often coupled with the use of "absolute" colorimetry, which generally refers to measured image colorimetry, absent any application of color-appearance transformations or media-relative colorimetric normalizations. Our experience is that many practitioners prefer this method of operation when a color-management system proves to be overly restrictive. Using absolute colorimetry, of course, effectively removes the color-management aspects of the system, but that is exactly what many users prefer. Doing so then allows them to implement their own methods of dealing with images, without the color-management limitations and interpretations imposed by the system itself. The system, of course, then becomes entirely dependent on tags. As discussed earlier, that approach does not directly solve the incompatibility problems associated with different types of imaging media, devices, and viewing environments. It does, however, provide the information necessary to subsequently do so using appropriate colorimetric transformations and color-encoding methods. In effect, then, the system simply becomes a means for color-image data transfer, and most color-management functions are performed using other applications.

Discussion

This chapter has described several types of colorimetric transformations, compared them to alternative transformations, and illustrated how the transforms we recommend can provide a practical implementation of the appearance-based color-encoding method of the Unified Paradigm. The recommended transforms are quite straightforward and well proven in numerous commercial applications. Consistent with the objectives of the Unified Paradigm, use of these transformations allows color to be represented unambiguously throughout an imaging system such that a desired color appearance is not subject to restrictions or limitations imposed by any particular interpretation, context, medium, or device.

Examples of the use of these transformations and the color-encoding method of the Unified Paradigm, alone and in combination with other color-encoding methods, are given in the two chapters that follow.

Summary of key issues

- Colorimetric transformations used for color-appearance encoding must account for viewing flare, image luminance level, observer chromatic adaptation, observer lateral-brightness adaptation, and observer general-brightness adaptation.
- Color encoding based on media-relative colorimetry will not properly account for observer general-brightness adaptation. Output images produced from such encoding will not have the correct brightness relationships.
- Encoding the color appearance of images from high-quality imaging devices and media requires the representation of luminance-factor values greater than those corresponding to the reproductions of perfect whites.
- In the CES of the Unified Paradigm, a clear and explicit distinction is made between an image illuminant and an observer's state of chromatic adaptation.
- Although the terminology used for many other color-management systems is similar to that of the CES of the Unified Paradigm, the respective encoding methods and underlying philosophies generally are fundamentally different. Such systems cannot provide the complete functionality of the Unified Paradigm.
- Systems having restricted CESs and that are entirely dependent on tagged image information are essentially color data transfer systems rather than color-management systems.

20

A Unified Paradigm: Example Systems

In this chapter, the Unified Paradigm and its associated color-encoding method will be explored further and demonstrated using a series of simple example systems in which various inputs, outputs, and image-interpretation options are selected.

The first task in designing each of these example systems, as well as any other color-imaging system based on the paradigm, is to define the reference viewing conditions for the color encoding specification. In practice, it might be advantageous to define these conditions differently for different systems. However, for illustration purposes, in this chapter we will specify a single set of encoding reference conditions. Doing so will allow us to focus our attention on the various types of colorimetric transformations required when the viewing conditions of actual inputs or outputs differ from those specified for the encoding reference.

In theory, encoding reference viewing conditions can be chosen essentially arbitrarily. It is more practical, however, to select conditions that are representative of those most likely to be used for actual input and/or output images. Doing so minimizes the magnitudes of the colorimetric transformations required to account for the appearance factors involved.

The encoding reference viewing conditions for the systems described here were chosen to be consistent with those typical of indoor viewing of reflection prints. In most respects, the conditions are also consistent with the viewing of back-illuminated transparencies and many types of electronic displays. The viewing conditions are defined in terms of the four characteristics that we consider to be most important:

- ***Viewing flare**:* Viewing-flare luminance is specified as equal to 1.0 % that of a white in the viewed image.
- ***Surround type**:* The surround is defined as average or normal, i.e., the area surrounding the image being viewed has a luminance factor of about 0.20 and is equal in chromaticity to the observer adaptive white.
- ***Luminance level**:* The luminance of a white in the viewed image is specified to be between 60 and 160 cd/m^2.
- ***Observer adaptive white**:* The chromaticity coordinates of the observer adaptive white are specified to be those of CIE Standard Illuminant D_{50}.

The examples that follow have been selected to demonstrate particular aspects of the Unified Paradigm and its associated color-encoding method.

Digital Color Management: Encoding Solutions, Second Edition Edward J. Giorgianni and Thomas E. Madden
 ISBN: 978-0-470-51244-9

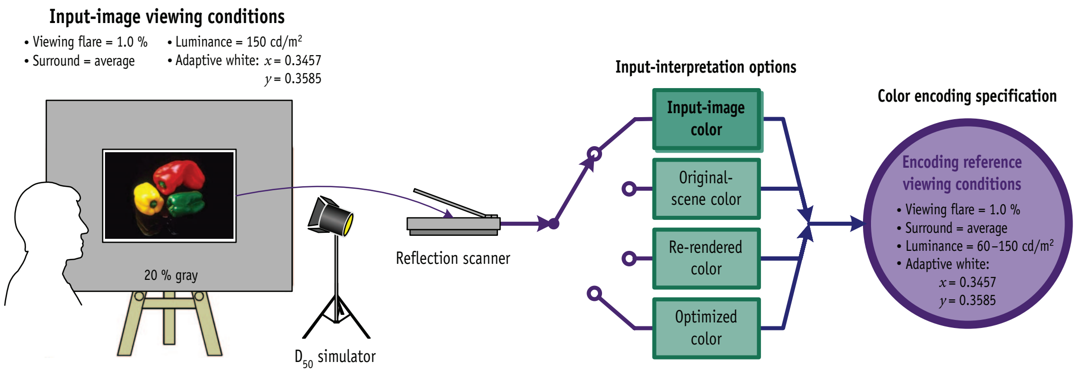


Figure 20.1a Input from a reflection image, as described in Example 1.

The specific color-appearance transformations described in the examples are quite simple, yet they have been shown to provide results that are quite satisfactory for most applications. Other, generally more complex, transformations could be used in cases where more demanding requirements must be met.

Example 1

In this first example, illustrated in Figure 20.1a, the image to be input to the system is a reflection print. The selected input-encoding option is to represent the color appearance of that print. The print is illuminated by a light source approximately simulating CIE Standard Illuminant D_{50}, and it is viewed under the following conditions:

- ***Viewing flare:*** Viewing-flare luminance, due to front-surface reflection from the print, is about 1.0 % that of a white in the print.
- ***Surround type:*** The print is surrounded by a neutral gray area of approximately 20 % reflectance. The surround type therefore can be considered average.
- ***Luminance level:*** The luminance of a white in the print is about 150 cd/m^2.
- ***Observer adaptive white:*** The observer is chromatically adapted to the chromaticity of the simulated D_{50} light source ($x = 0.3457$, $y = 0.3585$).

These actual input-image viewing conditions match the encoding reference viewing conditions in every respect; therefore, the reflection print can be encoded directly in terms of its standard CIE colorimetric values. These values would be based on the simulated D_{50} source. They could be measured using a colorimetric scanner or (with appropriate transformations) a densitometric scanner, as described in Chapter 11.

Figure 20.1b illustrates one procedure for developing and applying an input signal processing

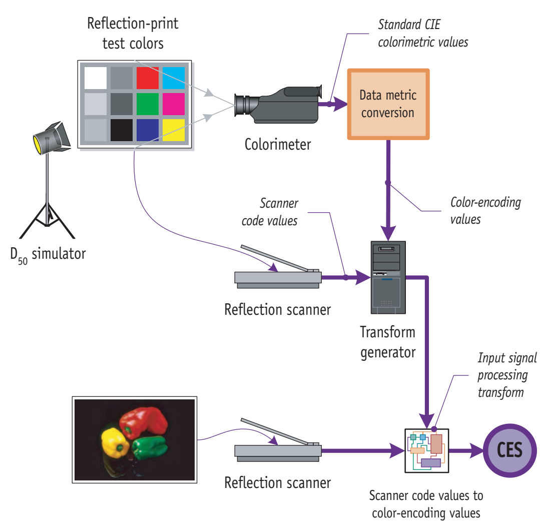


Figure 20.1b Derivation of a scanner input signal processing transform based on Example 1. No colorimetric transformations of the input image's standard colorimetric values are required because the actual input-image viewing conditions match the encoding reference viewing conditions.

transform based on this example. First, CIE colorimetric values are determined for an array of test colors on the reflection medium. These values can be computed from the spectral reflectances of the test colors and the spectral power distribution of the simulated D_{50} source, or they can be measured directly using a colorimeter and the light source. The colorimetric values then are converted to the particular data metric of the color encoding specification. The same test colors also are scanned to produce scanner code values. An appropriate application then is run to generate a mathematical transform relating scanner code values to color-encoding values. The resulting transformation, which may be in the form of one-dimensional lookup tables, matrices, polynomial equations, and/or three-dimensional lookup tables, then is applied in the input signal processing of the imaging system.

Example 2

The input in this example again is a reflection print (Figure 20.2a). The input-encoding option is to represent the appearance of that print when illuminated by a light source approximately simulating CIE Standard Illuminant D_{65}. The print is viewed under the following conditions:

- ***Viewing flare:*** Viewing-flare luminance, due to front-surface reflection from the print, is about 1.0 % that of a white in the print.
- ***Surround type:*** The print is surrounded by a neutral gray area of approximately 20 % reflectance. The surround type therefore can be considered average.
- ***Luminance level:*** The luminance of a white in the print is about 150 cd/m^2.
- ***Observer adaptive white:*** The observer is chromatically adapted to the chromaticity of the simulated D_{65} light source ($x = 0.3127$, $y = 0.3290$).

In this example, the actual input-image viewing conditions match the encoding reference viewing conditions except that the observer is chromatically adapted to the chromaticity of D_{65}, rather than to the chromaticity of D_{50}.

The reflection image can be encoded by first determining its standard CIE colorimetric values. These values would be determined using the spectral power distribution of the simulated D_{65} source. However, to be consistent with the reference viewing conditions defined for the color encoding, those colorimetric values must be transformed to visually equivalent colorimetric values for an observer who is chromatically adapted to the chromaticity of D_{50}. This can be done using any of a number of different chromatic adaptation transformation procedures, such as the von Kries procedure described earlier.

One result of the chromatic adaptation transformation is that an area of the image that appeared

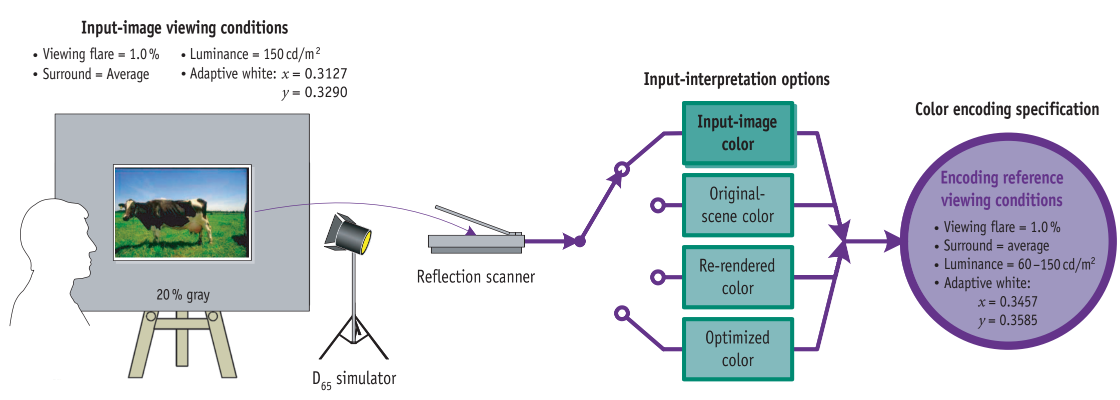


Figure 20.2a Input from a reflection image, as described in Example 2.

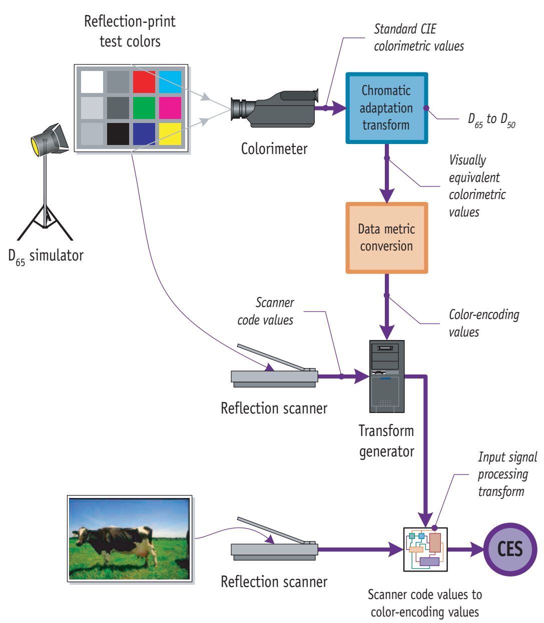


Figure 20.2b Derivation of a scanner input signal processing transform based on Example 2. The transform includes a transformation of measured D_{65} values to visually equivalent D_{50} values.

neutral (i.e., the area had a chromaticity equal to that of D_{65}) in the actual input-image viewing conditions would be encoded as a neutral for an observer adapted to the encoding reference viewing conditions (i.e., the area would have a chromaticity equal to that of D_{50}).

A scanner input signal processing transform for this example can be derived and applied using the procedures illustrated in Figure 20.2b.

The conditions of the next two examples admittedly are somewhat unusual, but they are not entirely unrealistic. They are similar to conditions used in art museums to display paintings, drawings, and photographs. The two examples have been included to illustrate the important (but somewhat elusive) distinction between the *actual input-image illuminant*, which directly contributes to the colorimetry of the input image, and the *observer adaptive white*, which describes the chromatic adaptation state of the observer who is viewing that image. As the examples show, the observer may not be chromatically adapted to the image illuminant.

Example 3

In the first of these examples (Figure 20.3a), the input is a reflection print that is illuminated by a small spotlight having a spectral power distribution that approximately simulates CIE Standard Illuminant D_{65}. The print is in a well-lighted room that is illuminated by a light source approximately simulating CIE Standard Illuminant D_{50}. The input-encoding option is to represent the color

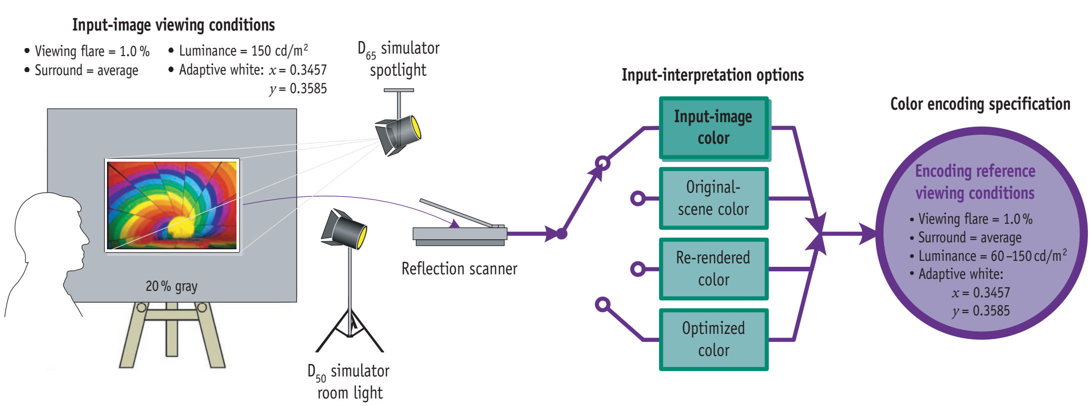


Figure 20.3a Input from a reflection print, as described in Example 3. Although the print is illuminated by a D_{65} spotlight, the observer is chromatically adapted to the chromaticity of the D_{50} room light source.

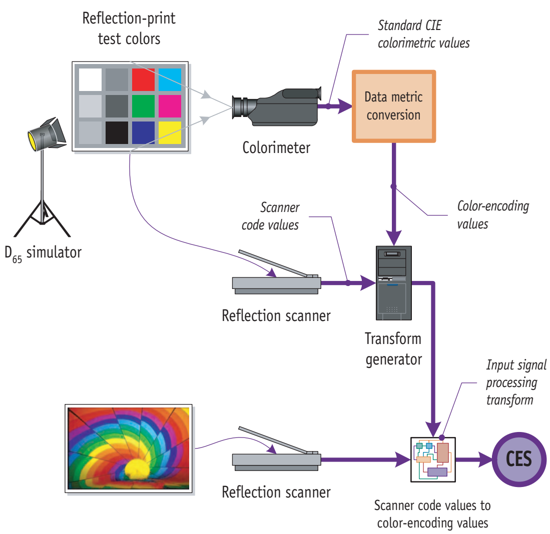


Figure 20.3b Derivation of a scanner input signal processing transform based on Example 3. No transformations of the colorimetric values of the input image are required because the actual input-image viewing conditions match the encoding reference viewing conditions.

appearance of that print. The viewing conditions are as follows:

- ***Viewing flare:*** Viewing-flare luminance, due to front-surface reflection from the print, is about 1.0 % that of a white in the print.
- ***Surround type:*** The print is surrounded by a neutral gray area of approximately 20 % reflectance. The surround area is illuminated by room light, but no room light falls on the print itself.
- ***Luminance level:*** The luminance of a white in the print is about 150 cd/m^2.
- ***Observer adaptive white:*** The observer is chromatically adapted to the chromaticity of the *room* illuminant (the simulated D_{50} light source).

As was discussed earlier, the observer adaptive white describes the chromatic adaptation state of the observer; it does *not* describe a reference viewing illuminant. Therefore, *the actual input-image viewing conditions still correspond to those of the encoding reference viewing conditions, despite the fact that the image itself is illuminated by a D_{65} light source*. The only difference from Example 1 is that the *colorimetry* of the print would be determined using the spectral power distribution of the D_{65} spotlight. That colorimetry, and therefore the color encoding, would reflect the fact that the print would appear somewhat cyan-bluish because it is illuminated by a D_{65} source and the observer is chromatically adapted to the (more yellow-red) chromaticity of D_{50}. A scanner input signal processing transform for this example can be derived and applied using the procedure illustrated in Figure 20.3b.

Example 4

In this example (Figure 20.4a), the input is a reflection print that is illuminated by a small spotlight having a spectral power distribution that approximates CIE Standard Illuminant D_{50}. The print is in a well-lighted room that is illuminated by a light source approximately simulating CIE Standard Illuminant D_{65}. The input-encoding option is to represent the actual color appearance of the print. The viewing conditions are as follows:

- ***Viewing flare:*** Viewing-flare luminance, due to front-surface reflection from the print, is about 1.0 % that of a white in the print.
- ***Surround type:*** The print is surrounded by a neutral gray area of approximately 20 % reflectance. The surround area is illuminated by room light, but no room light falls on the print itself.
- ***Luminance level:*** The luminance of a white in the print is about 150 cd/m^2.
- ***Observer adaptive white:*** The observer is chromatically adapted to the chromaticity of the room illuminant (the simulated D_{65} light source).

These viewing conditions *no longer correspond to the reference viewing conditions*, despite the fact that the print is illuminated by a D_{50} light source. The colorimetry of the input image should be determined using the spectral power of that D_{50} source. However,

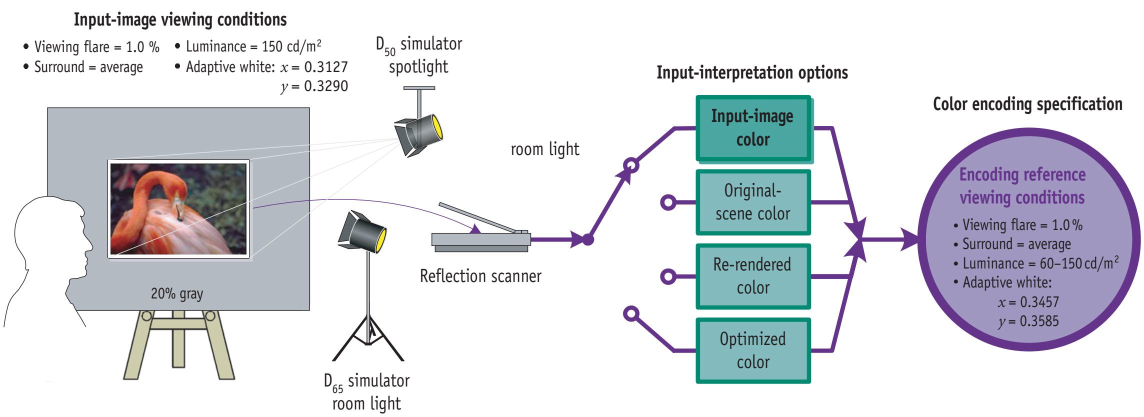


Figure 20.4a Input from a reflection image, as described in Example 4.

those colorimetric values then must be transformed, using a chromatic adaptation transformation from D_{65} to D_{50}, because the observer is adapted to the chromaticity of D_{65}. That transformation will reflect

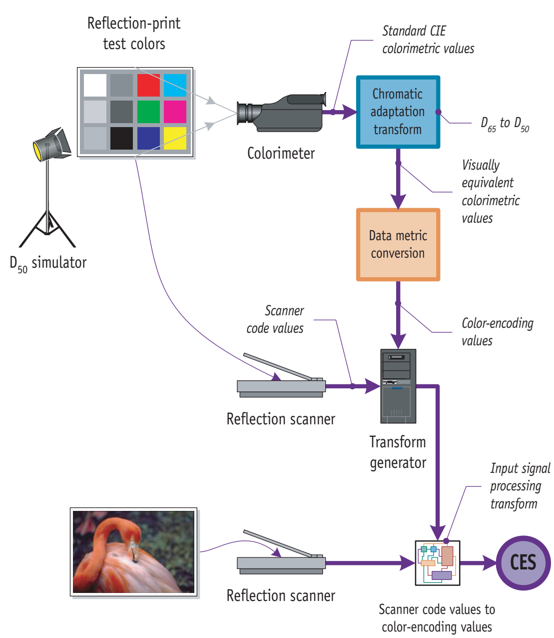


Figure 20.4b Derivation of a scanner input signal processing transform based on Example 4. A chromatic adaptation transformation is required here because the observer is chromatically adapted to the D_{65} room light.

the fact that the print would look somewhat reddish yellow because it is illuminated by a D_{50} source and the observer is chromatically adapted to the more cyan-blue chromaticity of D_{65}. An input signal processing transform for this example can be derived and applied using the procedure illustrated in Figure 20.4b.

Note that in both Example 3 and Example 4 it was stated that the observer is chromatically adapted to the chromaticity of the room illuminant. Depending on the specific conditions, the observer instead might be chromatically adapted to the chromaticity of the spotlight or to other chromaticities somewhere between those of the two illumination sources. Part of the "art" of color-appearance encoding is determining the chromaticity of the observer adaptive white. As described earlier, the criterion is that a stimulus having that chromaticity would appear perfectly achromatic to the observer.

Example 5

This example (Figures 20.5a and 20.5b) illustrates that the *difference* in the amount of flare in the input-image viewing conditions from that in the reference conditions must be accounted for in the encoding process. The input is a reflection print, illuminated by a light source simulating CIE Standard Illuminant D_{50}. The encoding option is to represent the

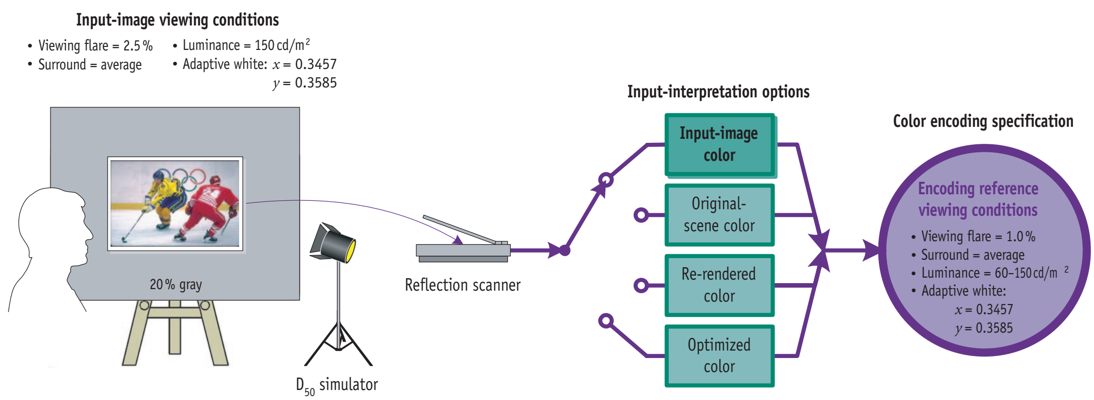


Figure 20.5a Input from a reflection image, as described in Example 5.

print's appearance when viewed under the following conditions:

- ***Viewing flare:*** Viewing-flare luminance, due to front-surface reflection from the print, is about 2.5 % that of a white in the print.
- ***Surround type:*** The print is surrounded by a neutral gray area of approximately 20 % reflectance. The surround type therefore can be considered average.
- ***Luminance level:*** The luminance of a white in the print is about 150 cd/m^2.
- ***Observer adaptive white:*** The observer is chromatically adapted to the chromaticity of the simulated D_{50} light source ($x = 0.3457$, $y = 0.3585$).

These conditions correspond to those of the encoding reference viewing conditions except that

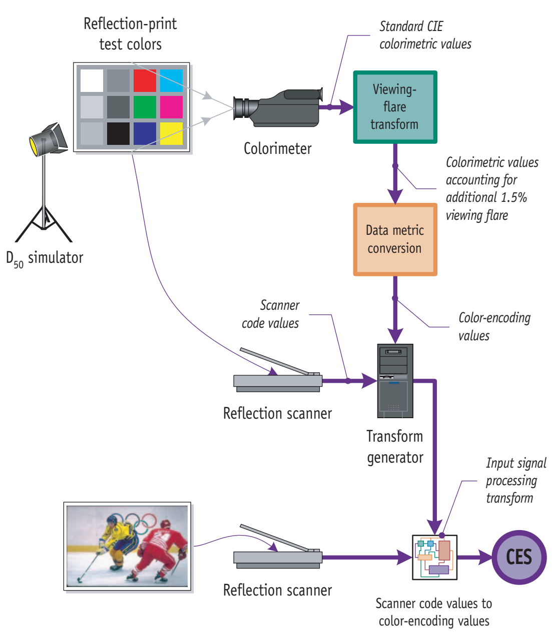


Figure 20.5b Derivation of a scanner input signal processing transform based on Example 5. The effect of the additional viewing flare must be accounted for in the input transform.

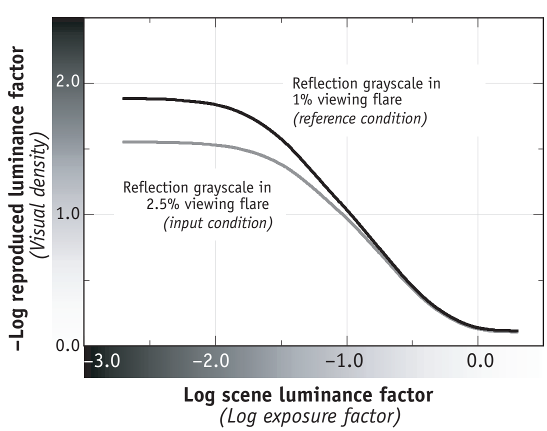


Figure 20.5c Comparison of reflection-image grayscales in viewing conditions of 1 % and 2.5 % viewing flare, as described in Example 5.

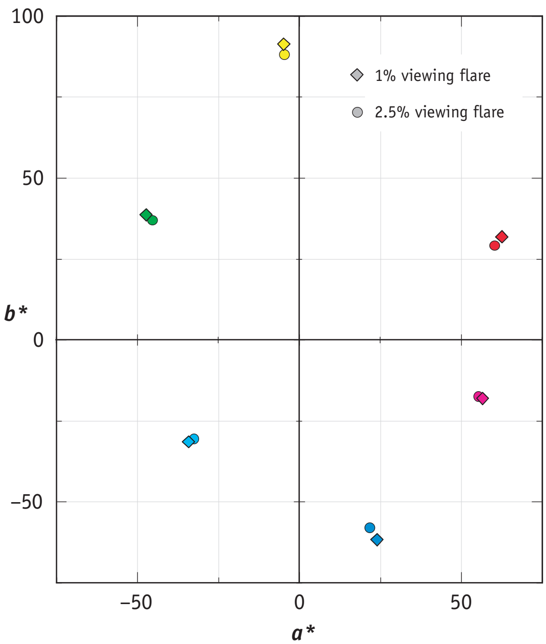


Figure 20.5d Comparison of CIELAB a^*, b^* values, reflection-print colors in viewing conditions of 1 % and 2.5 % viewing flare, as described in Example 5.

there is 2.5 %, rather than 1 %, viewing flare. The effect of the additional 1.5 % viewing flare must be accounted for in the encoding. This can be done using the transformation method described in Appendix E. The transformed colorimetric values will reflect the fact that in the actual input-image viewing conditions, the print will be lower in luminance contrast (shown in Figure 20.5c) and color saturation (Figure 20.5d) than it would be if viewed in the encoding reference viewing conditions.

Example 6

In this example, the input is a photographic negative. A direct encoding of the color appearance of that negative would make little sense (unless, for some reason, the user really wants backwards orange images). Appearance encoding of negative images requires some form of *rendering* of the scanned data.

An input signal processing transform for doing that can be constructed by first printing test negatives onto a suitable print medium, such as a photographic paper (Figure 20.6a). The colorimetry of the resulting reflection prints then can be measured and encoded using the same procedures described in the earlier examples of reflection-print input. In this particular example, the specified viewing conditions for the rendered image are those of the encoding reference conditions, so no additional colorimetric transformations are required. In cases where the specified viewing conditions for the rendered image differ from those of the encoding reference conditions, appropriate colorimetric transformations must be included in the input signal processing.

As an alternative to physically printing test negatives, data scanned from images on negative films can be "computationally printed" using mathematical models of various print media. For example, if the output medium is conventional photographic paper, the printing densities of the negative would first be computed, based on the effective spectral responsivities of that paper. Those printing-density values then would be used with a model of the paper to compute print colorimetric values.

This method is considerably more convenient than methods involving actual printing, especially when it is desired to produce transforms for a number of different input negative films as printed onto a variety of different print media. Moreover, the modeled print media used in these transformations need not be real. For example, because of their extended exposure and density dynamic ranges, the two hypothetical print media shown in Figure 20.6c each can render more information from a negative than can the actual photographic paper shown in the same figure. Use of such hypothetical media helps to minimize the loss of information that occurs when images scanned from negatives are encoded in terms of color appearance (Figure 20.6d).

Input signal processing transforms based on this example can be constructed, from either actual or computed prints, using the procedure illustrated in Figure 20.6b.

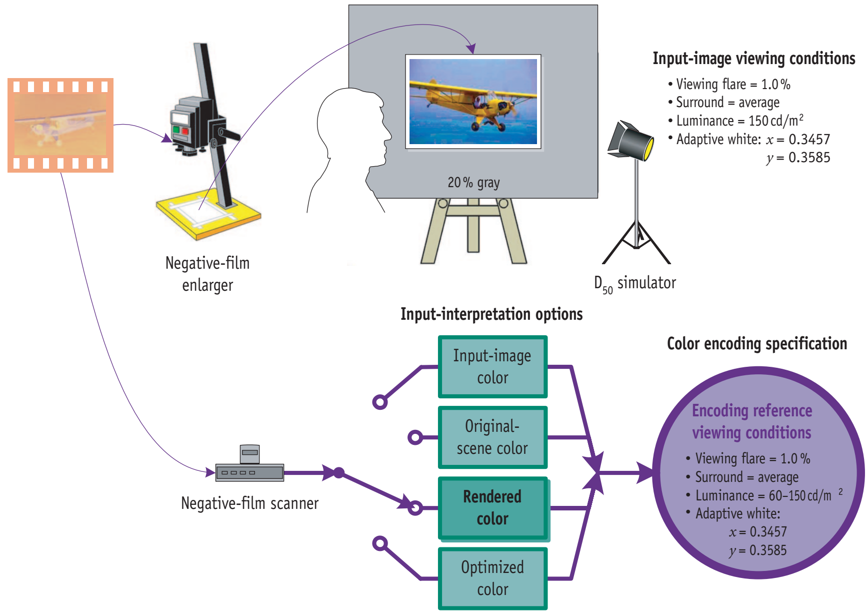


Figure 20.6a Input from a photographic negative, as described in Example 6. Data from a negative image must be rendered to a positive image for color-appearance encoding.

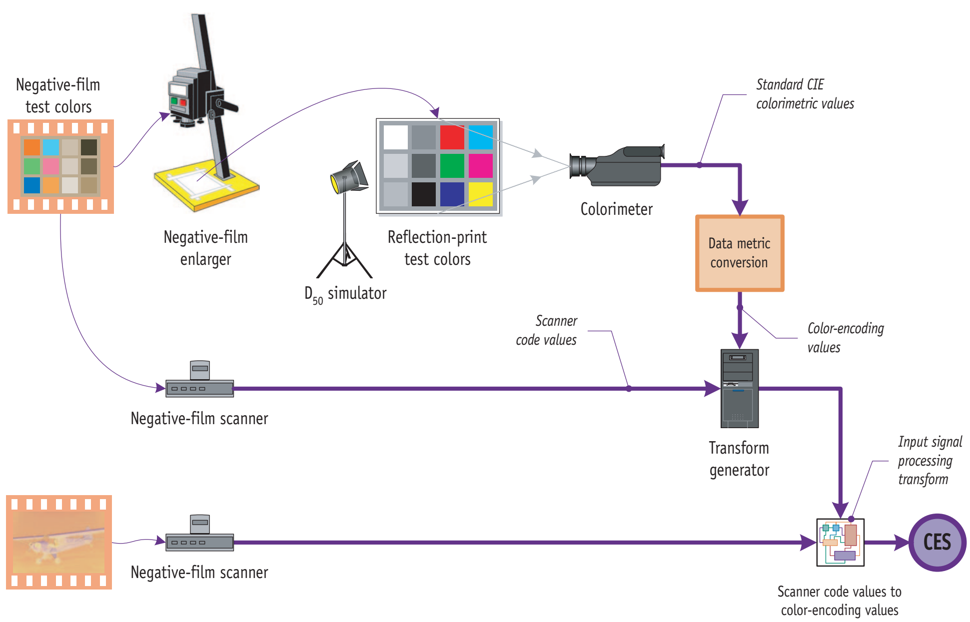


Figure 20.6b Derivation of a scanner input signal processing transform based on Example 6. Data scanned from negative images are rendered for color-appearance encoding.

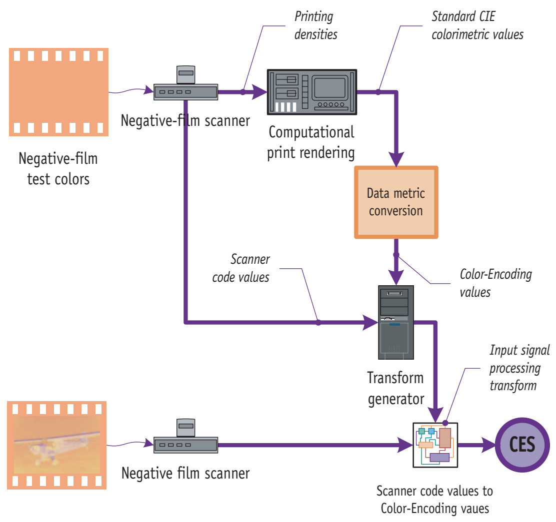


Figure 20.6c Grayscales for an actual photographic reflection-print medium and two hypothetical reflection-print media.

Example 7

The input for this example is a photographic slide (Figure 20.7a). The input-encoding option is to represent the appearance of the slide as projected in a darkened room, using a tungsten–halogen lamp. The slide is viewed under the following conditions:

- ***Viewing flare:*** Viewing-flare, due to stray projector light, equals about 1.0 % of the luminance of a white in the slide.
- ***Surround type:*** The areas surrounding the projected image are not illuminated. The surround type therefore can be considered dark.
- ***Luminance level:*** The luminance of a white in the projected image is 125 cd/m^2.
- ***Observer adaptive white:*** The observer is chromatically adapted to the chromaticities of the projected neutrals of the slide film. The average

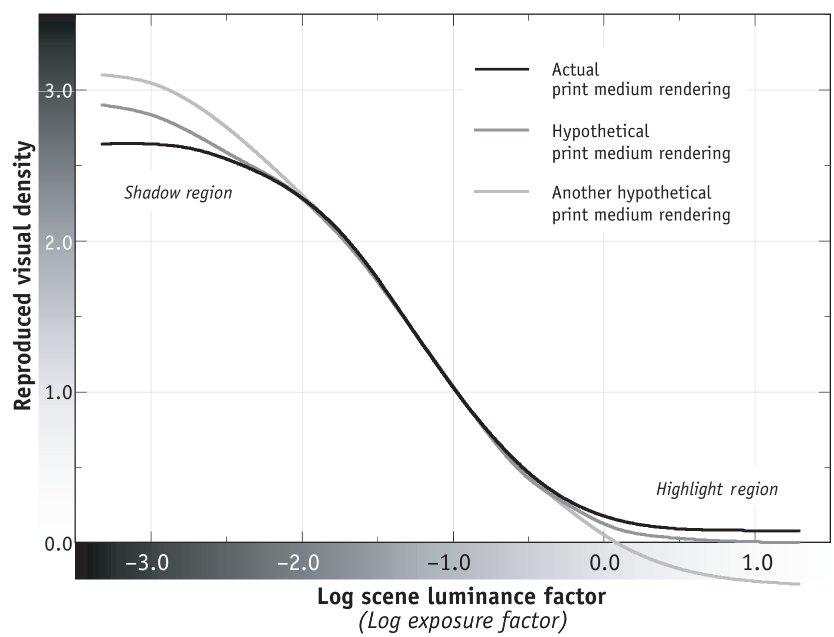


Figure 20.6d Rendering based on hypothetical print media helps minimize the loss of information that occurs when negative films are encoded in terms of color appearance. The renderings on the hypothetical media retain significantly greater information in the highlight and shadow areas.

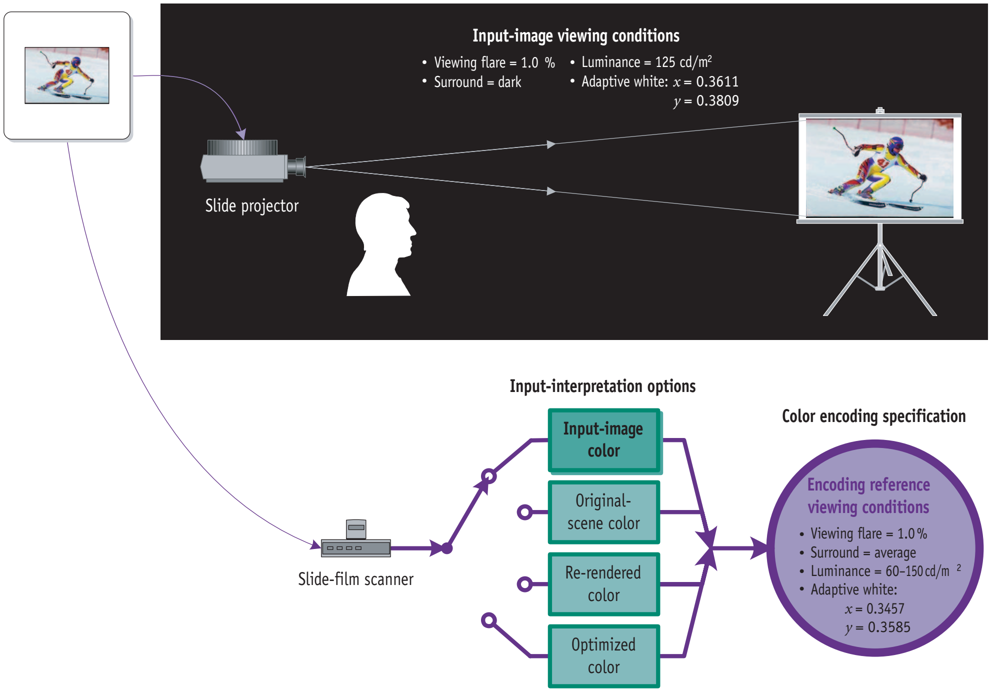


Figure 20.7a Input from a dark-projected slide, as described in Example 7.

chromaticity coordinates for the neutrals of this particular combination of film and light source are $x = 0.361$, $y = 0.381$.

In order to encode the appearance of the projected slide in terms of corresponding colorimetric values for the encoding reference viewing conditions, it is necessary to account for the effects of the dark-surround viewing environment. First, the reduction in perceived luminance contrast, due to observer lateral-brightness adaptation, must be considered. The encoded grayscale data must reflect the fact that a lower luminance-contrast image would be required in the reference conditions in order to match the appearance of the projected slide in the dark-surround conditions. The increase in brightness (lower apparent overall density, due to observer general-brightness adaptation) that occurs in the dark-surround viewing conditions also must be considered. Methodologies for determining the required colorimetric adjustments were discussed earlier, and they are described in greater detail in Appendix D.

In this example, it also is necessary to apply a chromatic adaptation transform to account for the fact that the observer is chromatically adapted to the average chromaticity of the projected neutrals of the slide film. Figure 20.7b illustrates the procedure for constructing an input signal processing transform appropriate for this example.

Example 8

The input in this example is a photographic slide taken of an original scene on a cloudy bright day (Figure 20.8a). The spectral power distribution of the daylight illumination corresponds to that of CIE Standard Illuminant D_{55}. The input-encoding option is to encode in terms of original-scene colors. The original-scene viewing conditions are as follows:

- ***Viewing flare***: There is no viewing flare in the original scene. This will be discussed in the example.
- ***Surround type***: The surround type is normal. This also will be discussed in the example.

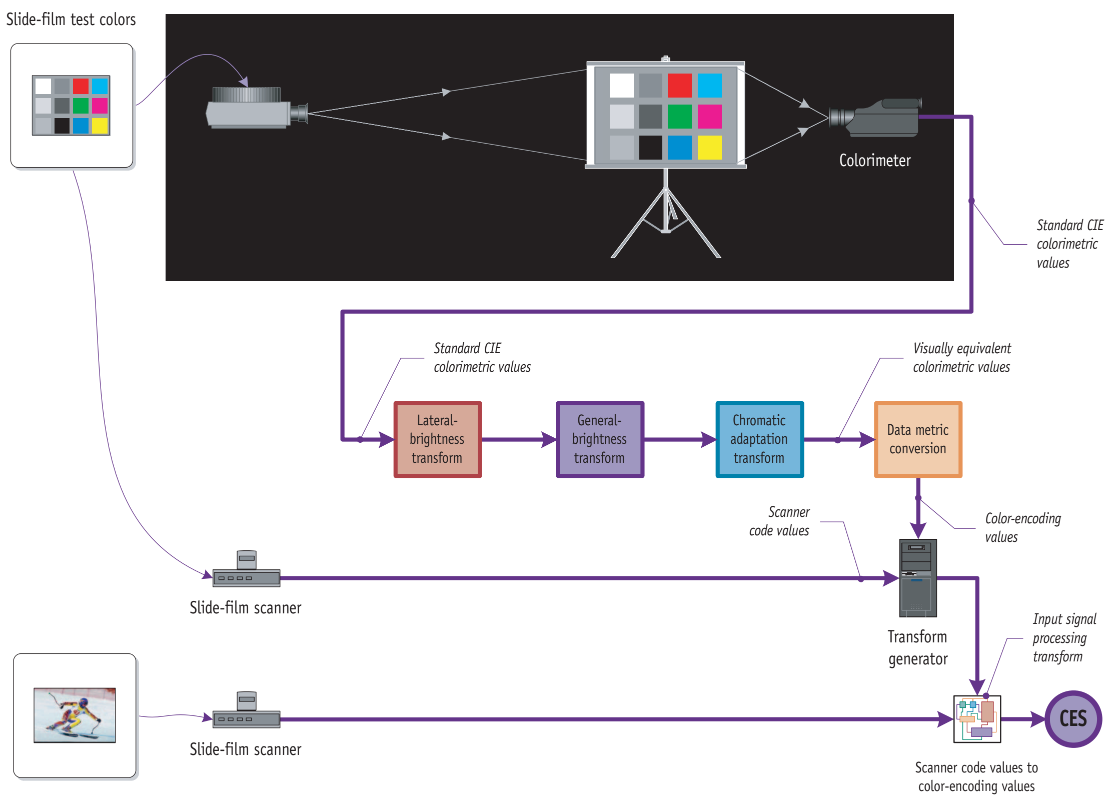


Figure 20.7b Derivation of a scanner input signal processing transform based on Example 7. The transform must account for the perceptual effects that result from the dark-surround viewing conditions. A chromatic adaptation transform is also required.

- ***Luminance level:*** The luminance of a white in the original scene is about 6000 cd/m^2.
- ***Observer adaptive white:*** The observer is chromatically adapted to the chromaticity of the D_{55} illumination ($x = 0.3324$, $y = 0.3474$).

Figure 20.8b illustrates a procedure for constructing an input signal processing transform based on this example. The transform relates scanner code values to *original-scene colorimetric values* that have been transformed to corresponding values for the encoding reference viewing conditions. Because the viewing conditions of the original scene differ significantly from the encoding reference conditions, several factors must be taken into account in this transformation.

First, it is necessary to incorporate a chromatic adaptation transform to account for the fact that the observer is chromatically adapted to the chromaticity of D_{55}. Next, colorimetric adjustments must be made to compensate for the reductions in perceived luminance contrast and colorfulness that result from the much lower luminance levels of stimuli in the encoding reference viewing conditions. Finally, compensation for the full 1 % viewing flare of the encoding reference viewing conditions also must be provided. This compensation is required because the original scene is considered to have no viewing flare. Any stray light present in the scene would appear as part of the scene itself. No adjustment is required for image surround. In an original scene, objects typically are surrounded by other similarly illuminated objects,

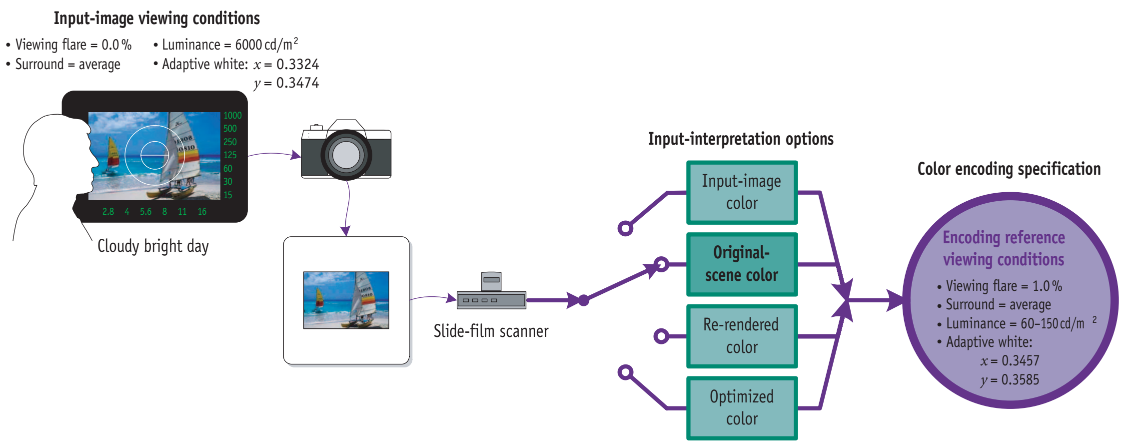


Figure 20.8a Color encoding of a photographic slide film in terms of original-scene colors, described in Example 8.

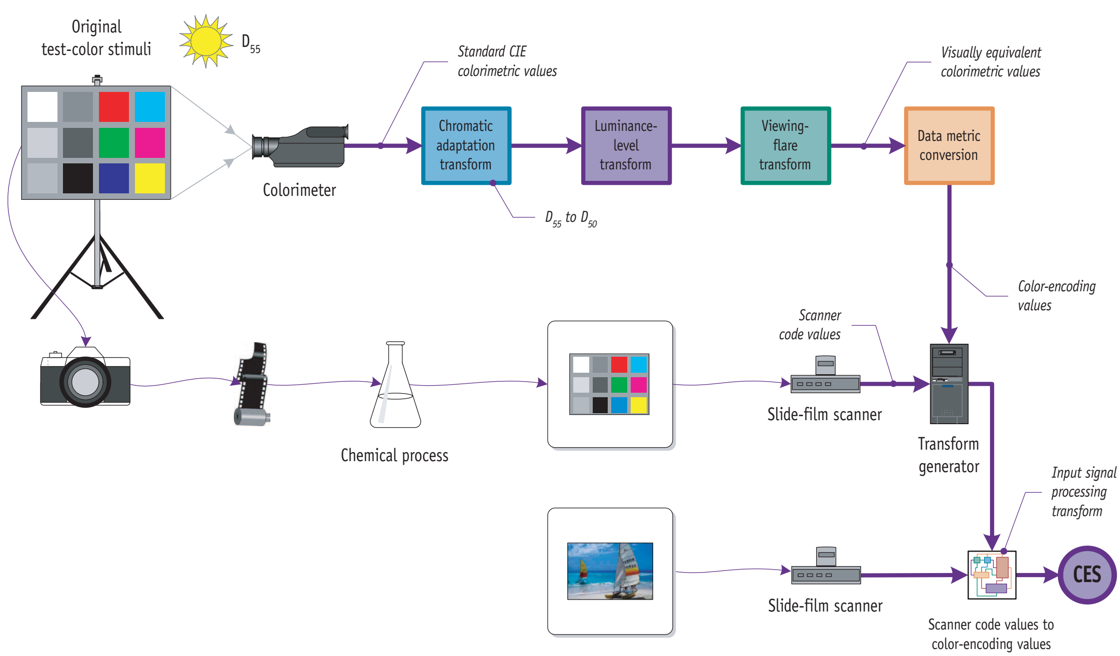


Figure 20.8b Derivation of a scanner input signal processing transform based on Example 8. Scene colorimetric values are transformed to corresponding colorimetric values for the encoding reference viewing conditions.

so the surround conditions can be considered to be average.

There are various ways that the required colorimetric adjustments of these transformations can be determined. For most applications, methods that are consistent with the properties of actual imaging products should be used. Doing so ensures input compatibility among images transformed from original-scene colorimetry to corresponding colorimetry, images computationally rendered from photographic negatives and from digital cameras, and reproduced images encoded in terms of their own appearance. Procedures for applying viewing flare, luminance contrast, and colorfulness compensations are described in Appendix G.

Example 9

The input in this example is a photographic slide (Figure 20.9a). The input-encoding option is to re-render the information scanned from the slide such that encoded values represent an image produced by a reference reflection-print system and viewed in the encoding reference viewing conditions.

Figure 20.9b illustrates a procedure for constructing an input signal processing transform for this example. Test colors are imaged using both the reference reflection-print system and the slide film. The colorimetry of the resulting reflection image then is determined and encoded in terms of the color encoding specification, using the techniques previously described for reflection images. The slide film is scanned, and a transform is derived relating scanner code values to encoded values for the corresponding colors of the reference reflection-print system.

Other re-rendering methods also can be used in this transformation. For example, the measured CIE *XYZ* *Y* tristimulus values for the slide film can be mapped to those of a reflection-print medium. The mapping can be determined based on the respective grayscale characteristics of the media. New *X* and *Z* tristimulus values then can be computed such that the x and *y* chromaticity values for the slide film are unchanged. This type of transformation, which is commonly used in prepress systems,

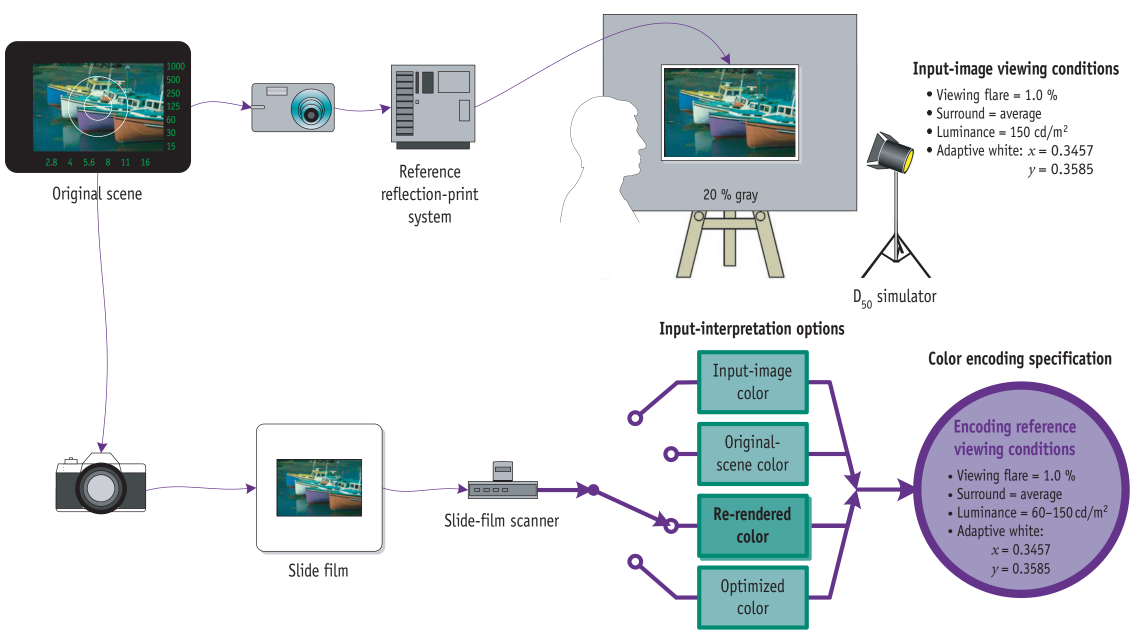


Figure 20.9a Input from a photographic slide, as described in Example 9.

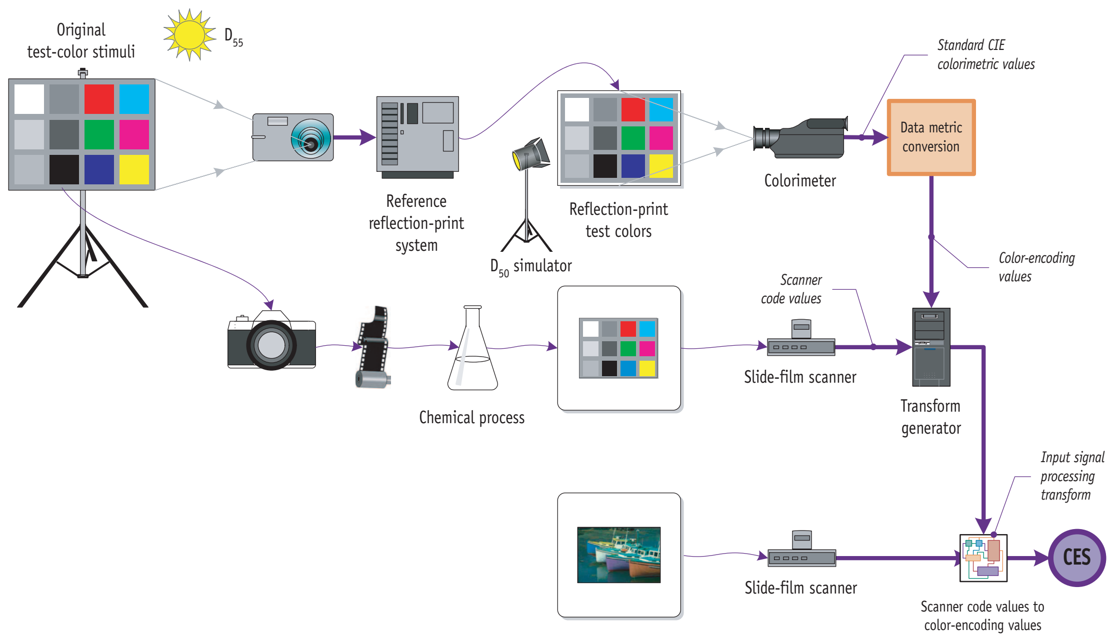


Figure 20.9b Derivation of a scanner input signal processing transform, based on Example 9. The transform re-renders the colors of the slide film to the colors that would have been produced by a reference reflection-print system.

retains somewhat more of the look of the slide film while producing a grayscale that is appropriate for reflection images. Additional colorimetric transformations may be required, depending on the viewing conditions specified for the re-rendered images.

Example 10

The input in this example (Figure 20.10a) is an image that has been created on a computer monitor. When used in this manner, a monitor is functioning as an *input* device rather than as an output display

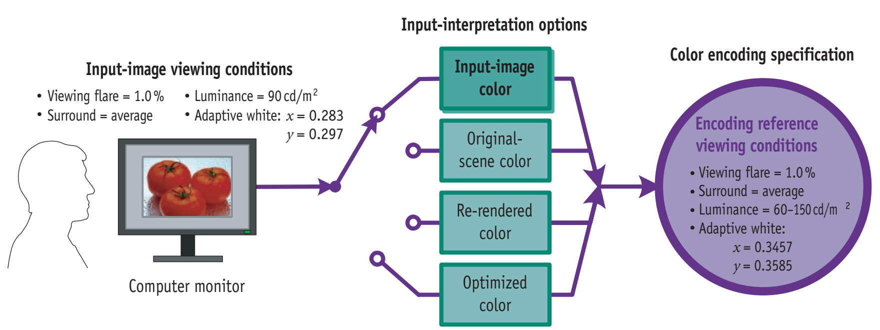


Figure 20.10a Input from a computer monitor, as described in Example 10.

device. In this example, the selected input-encoding option is to represent the appearance of the image as it appears on the monitor. The monitor display and the conditions under which it is viewed are as follows:

- ***Viewing flare:*** Viewing-flare luminance, due to reflections from the faceplate of the monitor, is about 1.0 % that of a white in the image.
- ***Surround type:*** The non-image area on the monitor is a uniform gray having the same chromaticity as the monitor white. The luminance of that gray area is approximately 18 cd/m^2. This value is about 20 % that of an image white. The surround type therefore can be considered average.
- ***Luminance level:*** The luminance of a white in the image is about 90 cd/m^2.
- ***Adaptive white:*** The observer is chromatically adapted to the chromaticity of the monitor white: $x = 0.283$, $y = 0.297$.

The actual input-image viewing conditions of this example match the encoding reference conditions quite closely. The only important exception is that the observer is chromatically adapted to the chromaticity of the monitor white, which is different from the D_{50} reference adaptive white.

The monitor image can be encoded by first determining its standard CIE colorimetric values. This can be done by direct measurement or, more practically, by the use of a mathematical model relating image *RGB* code values to colorimetric values for the resulting stimuli produced on the monitor. Because a monitor is an additive-color device, it is easily modeled, and the colorimetric values can be determined quite accurately. For color encoding, the measured or computed monitor colorimetric values must be transformed to visually equivalent colorimetric values for an observer chromatically adapted to the encoding reference adaptive white chromaticity ($x = 0.3457$, $y = 0.3585$).

Figure 20.10b illustrates a procedure for constructing an input signal processing transform for this example based on direct colorimetric measurements of monitor test colors. Figure 20.10c illustrates a comparable procedure in which colorimetric values instead are determined from a mathematical model of the monitor.

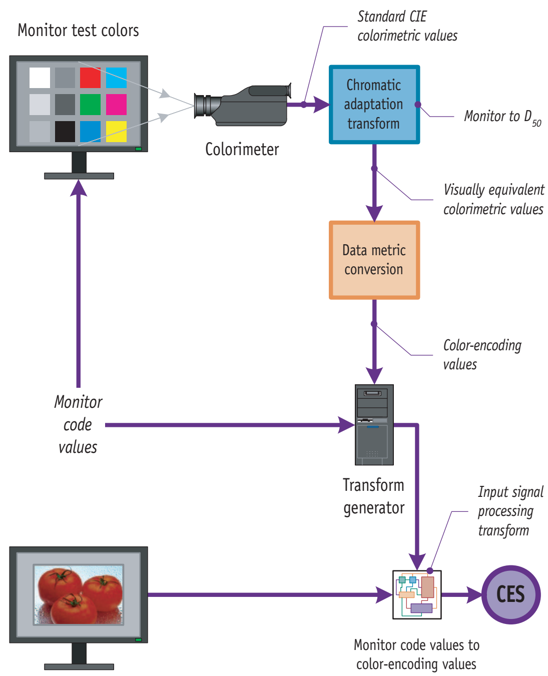


Figure 20.10b Derivation of an input signal processing transform for Example 10 based on direct colorimetric measurements of monitor test colors. The transform includes a chromatic adaptation transformation of monitor chromaticities to visually equivalent D_{50} chromaticities.

Example 11

In this example (Figure 20.11a), the input is from a digital still camera that has recorded the colorimetry of an original scene. The scene was illuminated by D_{55} daylight. The input-encoding option is to encode in terms of original-scene color. The viewing conditions for the scene are as follows:

- ***Viewing flare:*** There is no viewing flare in an original scene. As was previously discussed, any flare present in a live scene is considered part of the scene itself.

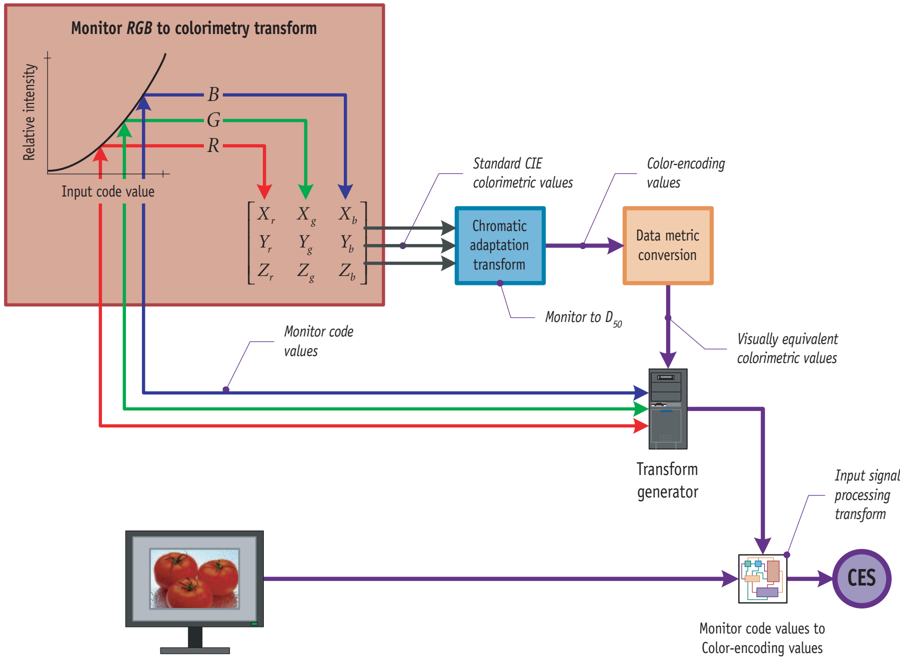


Figure 20.10c An alternative method for deriving an input signal processing transform for Example 10 using colorimetric values determined from a mathematical model of the monitor.

- ***Surround type:*** The surround type is normal. As previously discussed, scene objects typically are surrounded by other similar objects.
- ***Luminance level:*** The luminance of a white in the original scene is about 6000 cd/m^2.
- ***Observer adaptive white:*** The observer is chromatically adapted to the chromaticity of the scene's D_{55} illumination ($x = 0.3324$, $y = 0.3474$).

The viewing conditions of the original scene differ significantly from the encoding reference

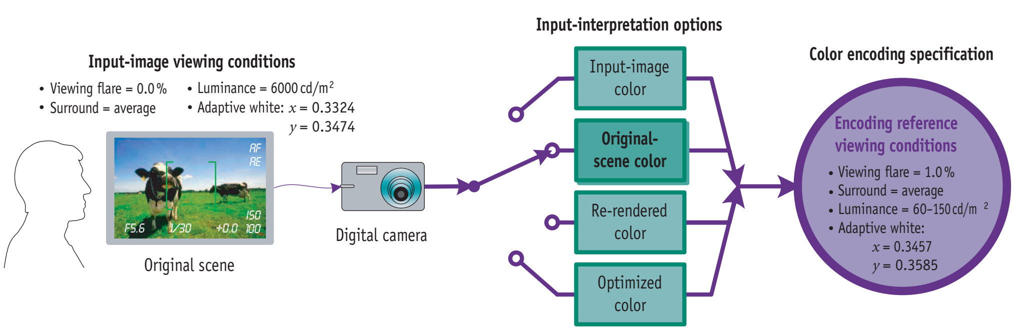


Figure 20.11a Input from a digital camera, as described in Example 11.

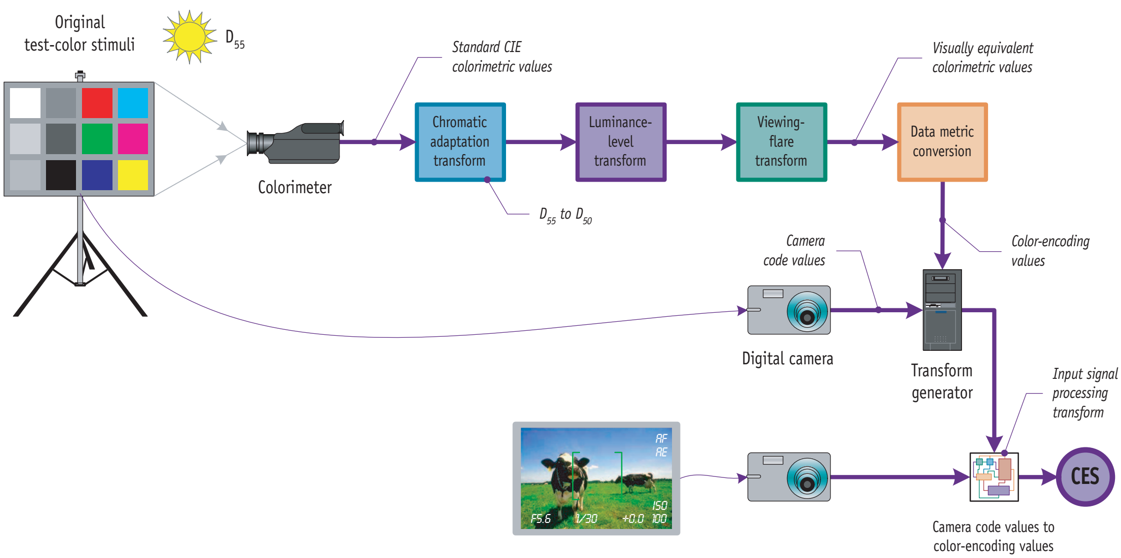


Figure 20.11b Derivation of an input signal processing transform, based on Example 11. The transform compensates for the effects of the lower luminance levels and greater flare of the encoding reference viewing conditions. A chromatic adaptation transformation (from D_{55} to D_{50}) also is required.

viewing conditions, so several factors must be taken into account, as shown in Figure 20.11b. A chromatic adaptation transform must be used, because the observer is chromatically adapted to the chromaticity of D_{55}. Colorimetric adjustments must be made to compensate for the reductions in perceived luminance contrast and colorfulness that result from the much lower stimuli luminance levels in the encoding reference viewing conditions. Compensation for the 1 % viewing flare of the encoding reference viewing conditions also must be provided.

For most applications, methods that are consistent with the properties of actual imaging products should be used in the transformation of original-scene colorimetric values to equivalent colorimetric values for the conditions of the encoding reference viewing environment. Use of such methods ensures that an appropriate degree of input compatibility is created for these images. This allows them to be used together with images that have been computationally rendered from digital cameras or photographic negatives and with reproduced images that have been encoded in terms of their own color appearance. The procedures used for the transformation of original-scene colorimetry to equivalent reference environment colorimetry are described in Appendix G.

Example 12

The final example of this chapter illustrates the use of output signal processing in the Unified Paradigm. This signal processing performs the three basic functions previously described in Chapter 14: colorimetric transformation, gamut adjustment, and code-value transformation. CES colorimetric values first are transformed by appropriate colorimetric transformations to produce visually equivalent colorimetric values for the actual output-image viewing environment. The methods used in deriving these transforms are the same as those used in the derivation of input colorimetric transforms. The directions of the transformations, of course, are opposite. Input transformations convert colorimetry from actual input-image viewing conditions to the encoding reference conditions, while output transformations convert colorimetry from the encoding reference conditions to the actual output-image viewing conditions that will be used.

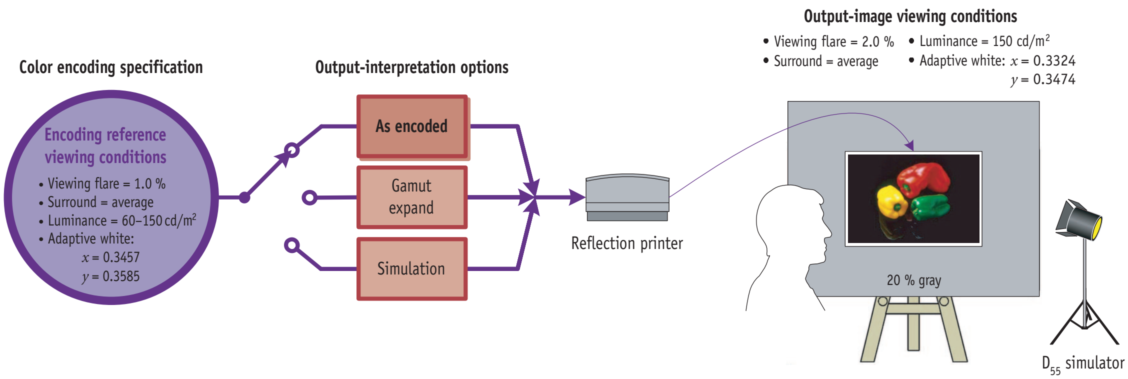


Figure 20.12a Output to a reflection printer, as described in Example 12.

The transformed colorimetric values define an output color stimulus that, when viewed in the actual output-image viewing conditions, will match the appearance of the encoded stimulus if it were viewed in the encoding reference viewing conditions. Because some of the resulting output color stimulus values may be outside the color gamut and/or luminance dynamic range of the output device or medium, some form of gamut adjustment (mapping) must be included in the output signal processing. Lastly, the gamut-adjusted values must be transformed to output-device code values, which can be done using the characterization and calibration methods described in Chapter 14.

In the example shown in Figure 20.12a, the output is to a reflection printer. The output-image option is to produce a displayed print that matches the color appearance described by the encoding. A source simulating CIE Standard Illuminant D_{55} will be used to illuminate the output print, which will be viewed in the following conditions:

- ***Viewing flare:*** Viewing-flare luminance, due to front-surface reflection from the print, will be about 2.0 % that of a white in the print.
- ***Surround type:*** The print will be surrounded by a neutral gray area of approximately 20 % reflectance.
- ***Luminance level:*** The luminance of a white in the print will be about 150 cd/m^2.

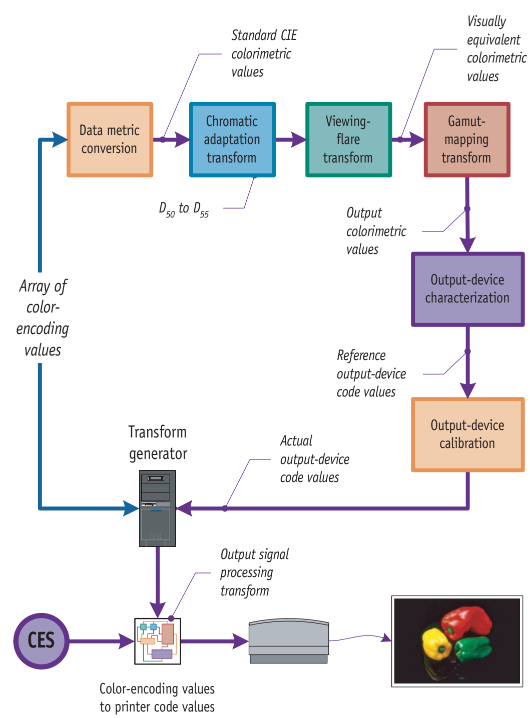


Figure 20.12b Derivation of an output signal processing transform based on Example 12. The colorimetric transform must include a chromatic adaptation transformation of encoded D_{50} values to visually equivalent D_{55} values and a compensation for the additional viewing flare of the output-image viewing environment.

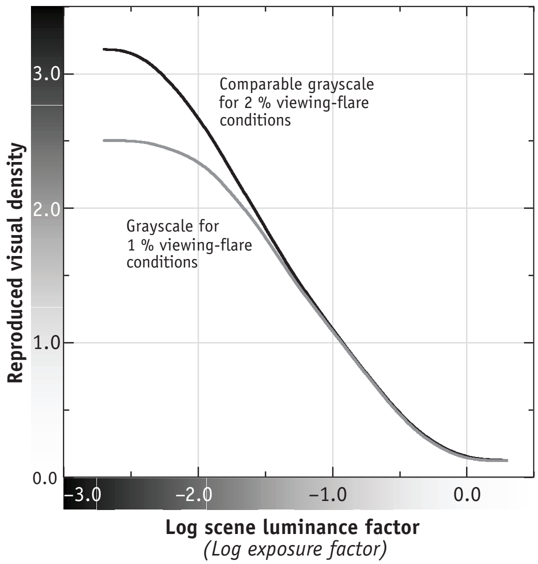


Figure 20.12c Comparison of reflection-image grayscales that will produce comparable color appearance when viewed in environments having 1 % and 2 % viewing flare, respectively.

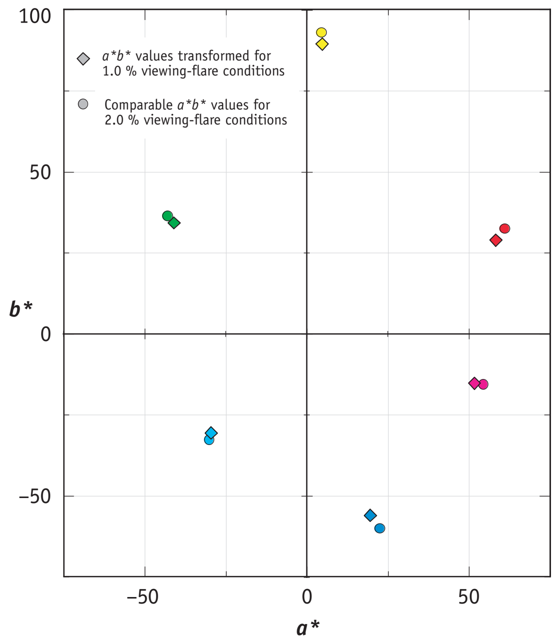


Figure 20.12d Comparison of reflection-image CIELAB a^*, b^* values that will have comparable color appearance when viewed in environments having 1 % and 2 % viewing flare, respectively.

- ***Adaptive white:*** The observer will be chromatically adapted to the chromaticity of the D_{55} source ($x = 0.3324$, $y = 0.3474$).

Because the observer will be chromatically adapted to the chromaticity of the D_{55} light source, encoded colorimetric values first must be transformed from encoding reference D_{50} colorimetric values to equivalent D_{55} colorimetric values. This can be done using the same type of chromatic adaptation procedure, such as a von Kries transformation, that was used for input encoding. Using the same method for input and output is operationally convenient, and it also results in an important overall symmetry of the system. So, for example, if the adaptive whites of both the input-image and output-image viewing environments are the same and all other factors are equal, the output colorimetric values will accurately match the input colorimetric values.

The effect of the *additional* 1 % viewing flare in the actual output-image viewing conditions, compared to that in the encoding reference viewing conditions, also must be accounted for in the output signal processing. The transformed colorimetric values must be higher in luminance contrast to help compensate for the additional flare (Figure 20.12c). The transformed values also must be higher in chroma in order to provide compensation for the loss in color saturation produced by the additional viewing flare of the actual output-image viewing conditions (Figure 20.12d). An output signal processing transform for this example can be derived using the procedure illustrated in Figure 20.12b.

Discussion

This series of examples has demonstrated that the use of appearance-based color encoding allows colors from various types of system inputs to be represented and encoded unambiguously, and without a restricted interpretation, for use by various types of system outputs. As a result of this flexibility, a diverse array of systems based on this encoding method can be made to function successfully according to any one of the *ABC* paradigms.

In the next chapter, we will show how the basic concepts of the Unified Paradigm and the principles illustrated by the examples in this chapter can be extended further to construct more complex systems involving multiple forms of input and output and multiple color-management paradigms.

Summary of key issues

- A comprehensive series of example color-imaging systems has demonstrated the basic technical approach of the unified color-management paradigm.
- Appearance-based color encoding was implemented in the systems by the use of a colorimetric specification and a defined set of encoding reference viewing conditions.
- Actual and encoding reference viewing conditions were specified in terms of viewing flare, luminance level, surround type, and observer adaptive white.
- In cases where the actual input-image viewing conditions matched the encoding reference conditions, input images were encoded directly in terms of standard CIE colorimetric values.
- In cases where the actual input-image viewing conditions differed from the encoding reference conditions, standard colorimetric values of input images were appropriately transformed for encoding.
- In cases where the actual output-image viewing conditions matched the encoding reference conditions, colorimetric transformations were not required in the output signal processing.
- In cases where the actual output-image viewing conditions differed from the encoding reference conditions, appropriate colorimetric transformations were included in the output signal processing.
- Gamut mapping and output-device characterization transformations also were included in the output signal processing.

21

A Unified Paradigm: Complex Systems

In this chapter, the flexibility and extensibility of the Unified Paradigm will be demonstrated in two example applications, both of which are considerably more complex than those studied in the preceding chapter. These examples will show that with the use of appropriate color encoding specifications, powerful systems can be assembled by linking simpler system components and subsystems in various parallel and serial arrangements. Also in this chapter the concept of *image states* will be introduced, and the principles and methodologies of scene-based color encoding and output rendering will be explored further.

A comprehensive color-imaging system

The requirements for the first example system, illustrated in Figure 21.1, might seem quite daunting because they include support for a diverse array of input and output media, devices, and image interpretations. On closer inspection, however, it can be seen that these specifications can be realized simply by combining the individual example systems previously described in Chapter 20 into a single system based on a common CES. Anticipation of this example was the reason an identical CES was used in all of the examples of that chapter.

As Figure 21.1 shows, the component inputs and outputs are both organized in parallel arrangements. With this strategy, images from any input medium, input device, and input-image interpretation can be transformed to the CES using appropriately selected input transformation methods, as described in Chapter 20. Likewise, CES images can be output to any medium, device, and output-image interpretation by the use of appropriate output transformations, which also were described in Chapter 20.

Some time ago, we and a group of colleagues designed and assembled a color-imaging system identical to that shown in this example. Our objectives were to further examine the concept of appearance-based color encoding and to determine its effectiveness in a practical, fully functional system. All combinations of input and output media, devices, and image interpretations were tested on the system, and all worked successfully.

It is important to emphasize that this success was not due simply to the diagrammatic arrangement of the various system components. Numerous other color-management schemes have been based on this type of arrangement, yet none has been capable of providing the full functionality specified for this complex system. What made our demonstration system work was a CES based on the color-encoding method described in Chapter 18. Use of this

Digital Color Management: Encoding Solutions, Second Edition Edward J. Giorgianni and Thomas E. Madden
 ISBN: 978-0-470-51244-9

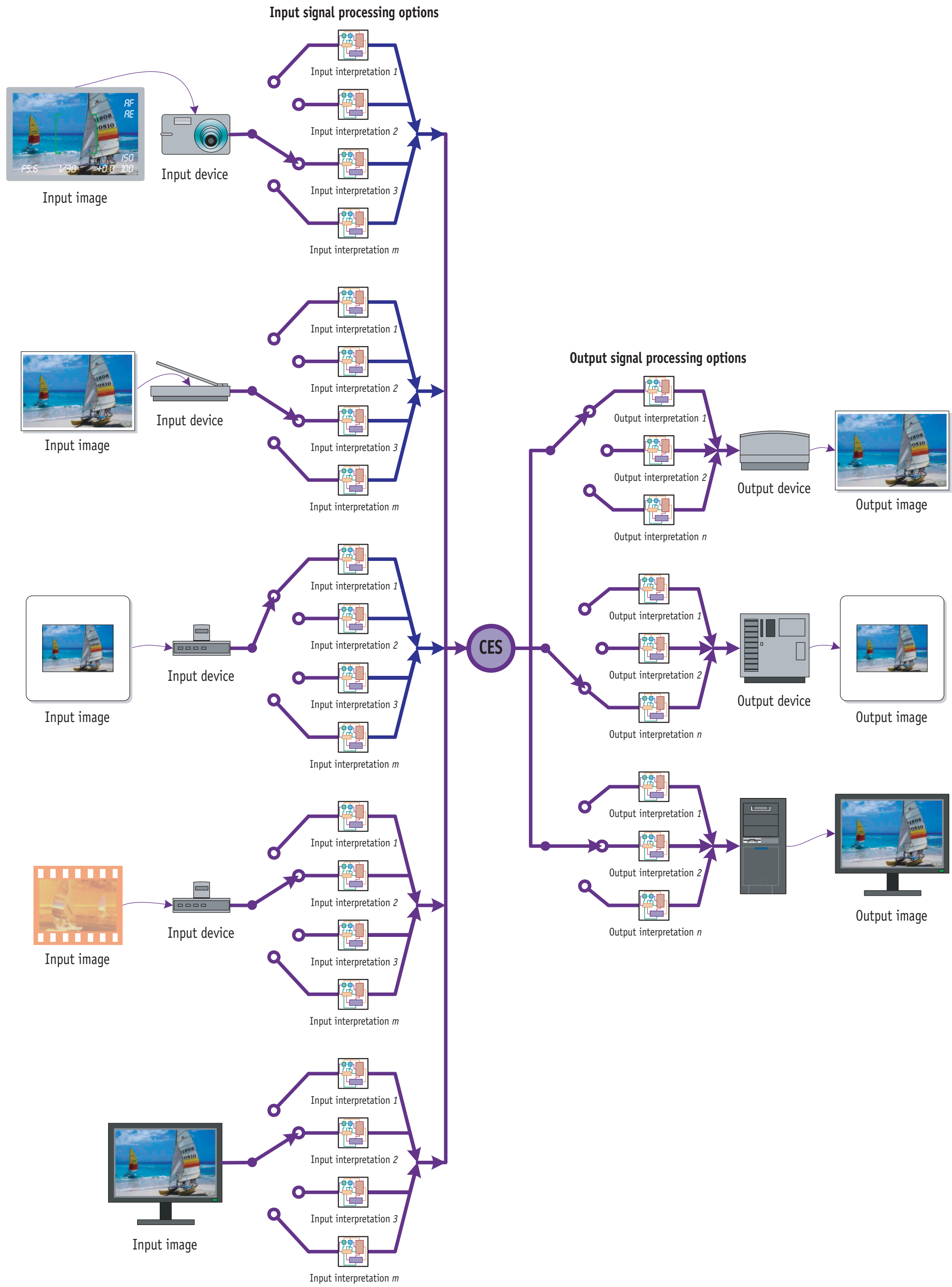


Figure 21.1 A complex multiple input, multiple output imaging system based on the Unified Paradigm. In order to meet the objectives defined for the system, its component inputs and outputs are organized in parallel arrangements.

color-encoding method, the color transformations described in Chapter 19, and an appropriate CES data metric provided the following system features:

- Images from disparate inputs could be encoded in the CES in a way that was both unrestricted and unambiguous.
- Images from disparate inputs could be edited and seamlessly composited in the CES without requiring knowledge of their origins.
- Output images could be produced such that their only restrictions were those unavoidably imposed by the physical properties of the output media and devices themselves.

It is particularly important to recognize that all input-image interpretations, all output-image interpretations, and all combinations of the two were fully supported by the system's CES. As a result, this single system was able to function according to *any* of the Type *A*, *B*, or *C* paradigms, as determined by the wishes of the user. Of course not every system needs that degree of flexibility, nor does every system need to support such a diverse assortment of inputs and outputs. However, the fact that a single system can do so demonstrates that, when well designed, color management itself need not impose any restrictions or limitations on how a system can be used.

In practice, most systems will function as subsets of this example system. If the methodology of the example is retained in those subsets, the only restrictions or limitations they will have will be those deliberately chosen by their designers. No additional restrictions or limitations will be imposed due to shortcomings in the paradigms, color management, or color encoding methodologies.

A digital cinema color-imaging system

In this second example, the Unified Paradigm will be applied to a particularly challenging undertaking—the design of a comprehensive system for digital cinema. The solution that will be described is based on fundamental concepts included in a proposal submitted by one of us (EJG) to the Academy of Motion Picture Arts and Sciences.

A comprehensive digital cinema system must support input of computer-generated imagery (CGI) as well as input of photographic images from digital cinema cameras, from conventional motion picture camera films, from conventional motion picture intermediate films, and from special photographic films designed exclusively for digital applications. In addition, the system must provide output to film recorders, digital projectors, various types of self-luminous displays, various types of digital storage media, and various devices for producing reflection prints and other forms of hardcopy images. The system described here is intended to meet all of these requirements and to be extensible in order to support image-capture and image-display devices and media that may be developed in the future.

As discussed in the first example and shown in Figure 21.1, the architecture of most color-management systems is based on the use of a single CES. In this classic arrangement, the data path from each of any number of inputs includes an input transform to convert input-specific signals to values specified by the CES. Similarly, the data path to each of any number of outputs includes an output transform to convert CES values to values that are specific to the particular output. When designed and implemented correctly, that architecture is sufficient to meet the requirements of most color-imaging systems, whether they are simple or as complex as the previous example.

The principal challenge in designing a given system based on that architecture is determining a CES that can support all the inputs and outputs while also meeting other system requirements for image quality and functions such as image manipulation and compositing. A logical place to start the design of the CES, then, is to determine the principal focus and functional requirements of the intended system.

Ordinarily, color-imaging systems fall into one of two categories: they are focused primarily either on the inputs (Type *A* systems) or on the outputs (Type *C* systems). For example, the principal focus of applications such as digital photofinishing is on the outputs. On the other hand, the emphasis of the Cineon digital film system, discussed in Chapter 10, was on the inputs. A principal objective of that system was to scan motion picture negatives in a way that retained all of the captured image information for use in image manipulation and for subsequent output to intermediate negative films. To support that function and to provide for the digital archiving of photographic negatives, an input-based CES defined in terms of photographic printing-density values was used. Even Type *B* systems, which are defined by a centralized color-encoding concept, still tend to favor either input or output. For example, most printing and publishing applications are based on the Type *B* paradigm, but their color encoding is intended to emulate output-image colorimetry.

What sets the proposed digital cinema system apart from the systems that have been discussed so far is that there are very demanding requirements that must be met for *both* the inputs and the outputs. Moreover, the input requirements are quite different from those of the output. As a consequence, a CES that meets one set of requirements will not meet the other, and a compromise CES that attempts to meet both will be satisfactory for neither.

In particular, the full dynamic range of acquired input-image information cannot be retained in a CES that is also appropriate for display by output devices and media. That alone might suggest the use of an input-based CES. However, it is also essential for the system to include a means for unambiguously specifying and representing the intended output appearance of the final display. What is required, then, is an alternative architecture that allows *all* encoding-related objectives for this complex system to be achieved without compromise. Perhaps the best way to explain the principle of that architecture is to first introduce a concept that the authors refer to as *image states*.

Image states

The term "image state" arose from a discussion we had with a colleague—a computer programmer—regarding the fundamental difference between scene-based color encoding and other forms of encoding. Our colleague said that it is common in programming to describe the flow of data as a series of different *data states* created by various state-to-state *transitions*. She asked if similar terminology could be applied to the flow of image data. We thought that was an excellent concept that might help explain that there are operations in an image signal processing chain that do more than simply *adjust* the colorimetry of an image; they result in a *transition* of the image to a different colorimetric *state*.

As an analogy, consider that water can be adjusted in temperature over a fairly wide range without changing its properties in a fundamental way. However, liquid water also can transition to a completely different state, i.e., to a solid, ice, or to a vapor, steam. It is useful to think of images in the same way. They can exist in terms of a scene-space encoding, where they can readily be edited and adjusted without changing their fundamental nature. We can say, then, that they are in a *scene state*. In that state, their colorimetry is not appropriate for direct viewing in typical display environments. In order to produce image colorimetry for display, scene-state colorimetry cannot simply be adjusted. Instead, it must be *transformed* (or *transitioned*) to a new state that we can refer to as a *rendered state*.

To our dismay, shortly after the concept of image states was introduced to the industry, others attempted to take the idea further by defining various types and subsets of scene states and rendered states. Some even suggested that intermediary image values, such as printing densities or video code values, should be classified as unique states. Various image-state diagrams then were constructed in an attempt to clarify the complex interrelationships among the various proposed states. Later, still more states were defined including "scene referred," "focal-plane referred," "reference-input-medium referred," "code-value referred," "darkroom referred," "output referred," "reference-output-medium referred," "picture-referred," and "indirectly output referred." Although these efforts no doubt were made with the best of intentions, our opinion is that they only served to make things far more complicated than necessary. Moreover, the fundamental distinctions between ordinary image adjustments, straightforward data metric conversions, and true image-state transitions were lost. In this discussion, then, we will maintain the original concept of just two fundamental image states: *scene state* and *rendered state*.

Recognition of the distinction between these two states is essential to understanding the proposed architecture of the digital cinema system. Both states have their appropriate uses, and in a digital cinema system, both are required. As a consequence, a system architecture that incorporates and fully supports *both* states is necessary.

Digital cinema system architecture

The alternative architecture, shown in Figure 21.2, is designed to meet all objectives of the digital cinema system. In the multiple input/output arrangement shown in the figure, the inputs and outputs have different CESs, each of which is optimized to meet a different set of criteria. Although this architecture might appear to be a significant departure from the Unified Paradigm, it is not. Instead, it is a demonstration of the flexibility and extensibility of that paradigm. As in many previous examples, the architectural arrangement includes multiple input and output transforms arranged in *parallel* to provide support for various inputs and outputs. In addition, however, the architecture also includes a *series* arrangement of components. It effectively combines two Unified Paradigm systems, where the output of the first component system becomes the input to the second.

The first component system is based on the Type *A* paradigm. It uses an input CES (ICES) that supports scene-state images. The ICES is designed to encode all information that can be acquired by existing and anticipated future input devices and media, including those having the greatest expected image-capture capabilities. Input images are transformed to the ICES through the use of appropriate input device transforms (IDTs).

The second component system is based on the Type *B* and *C* paradigms and uses an output CES (OCES) that supports rendered-state images. The OCES is designed to encode the color appearance of output images, rendered within the system from ICES images, as displayed on a hypothetical display device having both a luminance dynamic range and

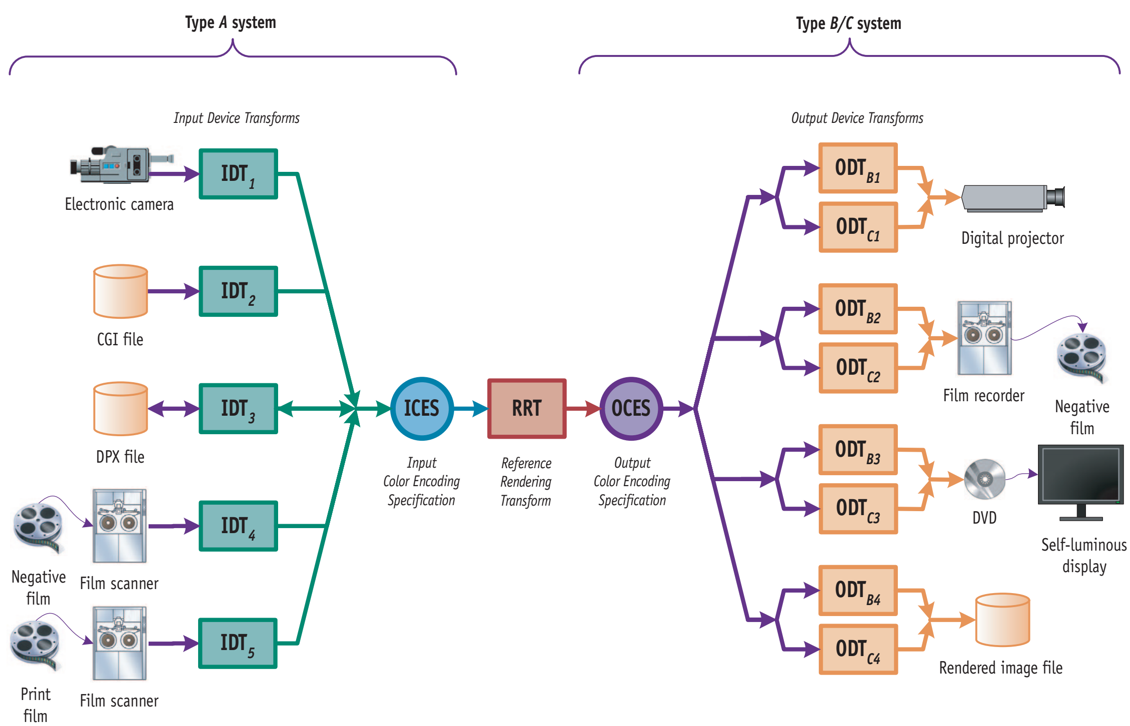


Figure 21.2 This system architecture, designed for digital cinema, includes both parallel and serial arrangements of components. The inputs and outputs have different color encoding specifications, each optimized to meet different requirements.

color gamut that meet or exceed those of present or anticipated display devices and media. Following the OCES are multiple output device transforms (ODTs) that are used to transform OCES values to device-specific colorimetry and device-specific code values for the various system outputs. In some applications, the output objective may be that all outputs match the color appearance specified in the OCES (and thus match each other) as closely as possible—the Type *B* paradigm. For other applications, the objective instead may be for each output to produce the best images it possibly can, given its particular capabilities and limitations, with less emphasis on matching images from other types of outputs—the Type *C* paradigm. Note that the option to choose between these output paradigms is included in the system diagram (Figure 21.2).

Connecting the two CESs (ICES and OCES) of the two component systems is a necessary image-state transition that we refer to as a reference rendering transform (RRT). The RRT transforms the colorimetry of scene-state input images to colorimetry appropriate for the reference display device and its associated reference viewing environment. Details regarding each CES, their associated input and output transforms, and the connecting RRT are discussed in the sections that follow.

ICES encoding method

The color-encoding method for the ICES of the system is that of scene-based color encoding, which was introduced in Chapter 12. In the most basic form of this encoding method, original-scene color stimuli remain in scene-state space and are represented by their original colorimetry rather than by values corresponding to those of intermediate signals or to those of rendered reproductions. There are many properties inherent in this method of encoding color information that make it well suited for the digital cinema ICES. In particular, it supports input from any and all types of media and devices, regardless of their disparity, and it allows information to be encoded in a way that places no limits on luminance dynamic range or color gamut. Any color that can be seen by a human observer can be represented by this method. (In some cases it may be appropriate to limit the mathematical scale of the encoded space in the data metric, but that is a separate design issue. Scene space itself is without such limits.)

By contrast, any of the alternative encoding methods based on rendered-state spaces will, essentially by definition, have restricted luminance dynamic ranges and limited color-gamut boundaries. As a result, some image color information captured in acquisition would be lost in an encoding process based on rendered-state spaces. The extent of that loss would depend on exactly how the space is defined. However, if the ICES were based on the rendering properties of any real display device or medium, the loss would be considerable. This would be unacceptable for image adjustment, for the addition of visual effects, and for other types of image manipulation that are performed in the earlier stages of motion picture post-production where the greatest flexibility is required.

Referring again to Figure 21.2, it is important to note that as in any imaging system, the flow of information in this system is "downhill," i.e., less and less information is passed as image data flow from the first block to the last. The most color information available is that which is originally provided by the input. If the input transformations are implemented correctly and the ICES is designed well, no useful information will be lost in the input encoding process. This is important in motion picture work because exactly what information will be needed generally is not known until all image manipulations have been finalized.

In this system, then, the underlying strategy is to retain as much information as possible for as long as possible. This means that color information should remain in the ICES, or in an equivalent scene-based workspace, until all creative work on an image is complete and the image is ready for rendering to output. A similar strategy has been used for decades in film-based cinema systems. Scenes are recorded on a long-latitude camera negative film, and post-production work is performed using the original negative or a duplicate made using an intermediate negative film. It is only after all such work is complete that images are rendered for output using a photographic print film specifically designed for theater projection. A comparable strategy is appropriate for digital cinema.

Color in scene-based encoding

Before discussing how ICES values can be produced from various types of inputs, it is important to address a common misconception regarding the meaning of color in scene-based encoding. It was just stated that, in its most basic form, scene-based encoding represents original-scene color stimuli in terms of their actual colorimetric values. That unique capability is ideal for some imaging applications. For example, an accurate representation of original-scene colorimetry is a necessary first step when colorimetric exactness is the principal objective of the entire imaging system. Scene-based encoding therefore is used in many scientific, medical, law enforcement, military, and other technical applications. It is also appropriate in more conventional applications where accurate color reproduction is important, such as in the imaging of paintings and other art objects for high-quality printing and other forms of display.

One possible concern regarding the use of scene-space encoding for digital cinema might be that there are many situations where the color objective is *not* one of accuracy. Instead, the desire might be to deliberately alter colors. For example, a cinematographer might select a particular film or electronic camera because it "distorts" scene colors in a way that helps create an artistically desirable look. Similarly, a colorist may want to digitally alter encoded scene colors or a CGI artist may wish to create entirely unrealistic scene colors to represent an alien world.

These concerns clearly are legitimate for motion picture production. However, they are based on a common assumption that encoded colorimetric values are in scene space *only* if they accurately represent actual scene color appearance. That assumption is incorrect. It results from confusing the *values* encoded in a scene space with the properties of the space itself. It is very important, then, to understand the following: *The use of scene-based color encoding is not limited to the representation of actual scene color appearance, either under a set of reference conditions or under any other set of conditions.*

That statement might sound contradictory, but consider this example. Imagine an outdoor scene on a heavily overcast day. Now imagine the same scene, but on a clear sunny day. Under the sunny condition, the colors would be much more saturated. So what, then, are the "original colors of the scene?" The answer is that they are not fixed because they can be altered by variables such as lighting conditions. Since it would be perfectly valid to accurately record the scene colorimetry under the conditions of either day, it is equally valid to digitally alter one day's image to look like the other. In fact, it is perfectly legitimate in scene space to alter the color of any image for any reason. It is also legitimate to use CGI techniques to create "virtual scenes" having *any* desired color properties, whether they are accurately realistic or completely imaginary. Scene space, then, can incorporate scene colors that actually existed, colors that were altered by lighting or exposure settings, colors that were altered by the characteristics of an input device or medium, colors that were modified by a colorist, and colors that were created entirely from imagination.

The somewhat nebulous meaning of color in scene-based encoding has been a source of some concern and considerable confusion since we first introduced the method to the industry years ago. Because this subject is so often misunderstood we will have more to say about it throughout the remainder of this chapter. For now, it may be helpful to consider this fact regarding scene-based encoding in general and the ICES of this example system in particular: scene-based encoding provides an appropriate means for unambiguously specifying the appearance of color *in the ICES itself*, but it does *not* specify *how*, nor does it specify *why*, a given ICES color came to be. That is left entirely up to the system user.

Getting into ICES space

In the following series of input examples, the process of transforming images from a number of different types of input devices, media, and image files into ICES scene space will be discussed in detail. The encoding method of the ICES will be based on scene-space CIE colorimetric values. In practice, various color primaries and color spaces could be used to express that colorimetry, but such decisions are data metric issues and not encoding-method issues. Therefore we will simply say that the objective of the input transforms in the examples is to produce

ICES values expressed in terms of CIE *XYZ* tristimulus values, with the understanding that these values instead could be expressed in terms of another data metric if desired.

Also, the input process will be simplified in these input examples by temporarily making the assumption that the image-taking conditions are the same as the ICES reference viewing conditions. This will allow us to concentrate on other aspects of the input-encoding process. Issues regarding the relationship of actual input-taking conditions and reference input-encoding conditions will be covered later in this discussion.

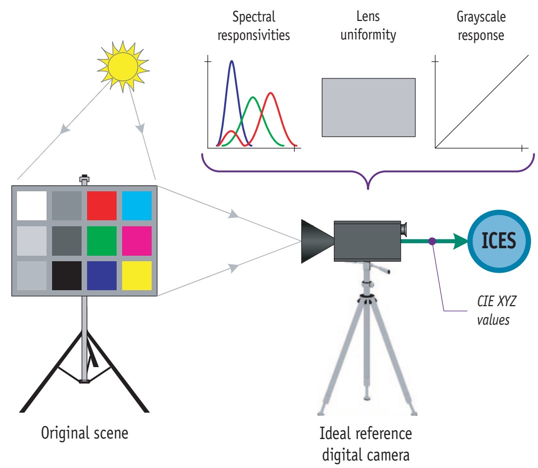


Figure 21.3 A reference camera being used to capture the color stimuli of an original scene. No input transform is required because the code values produced by the reference camera directly correspond to CIE *XYZ* values, as specified by the ICES.

ICES from an ideal digital camera

The most direct input device for scene-based encoding would be an ideal reference digital cinema camera having the following properties:

- The spectral responsivities would exactly match those defined for the CIE 1931 Standard Observer.
- The amount of camera flare would be representative of that present in actual motion picture cameras.
- Other optical characteristics would be perfect. In particular, there would be no spatial nonuniformity.
- The relationship between sensor exposure levels and digital code values output by the camera would be perfectly linear.

Figure 21.3 illustrates a reference camera being used to capture the color stimuli of an original scene. Note that no IDT is required for this device because the code values it produces directly correspond to CIE *XYZ* values, as specified for the ICES.

This first example is straightforward, but it is also important because it illustrates the fundamental concept of scene-based color encoding. Moreover, it provides a point of reference for other input sources. The examples that follow will discuss various inputs that depart in one or more ways from this idealized reference input. In each case, an input-specific IDT will be required to compensate appropriately for such departures. The IDTs could, for example, be designed such that ICES values derived from any of the inputs match as closely as possible those that would have been directly generated by the reference camera had it captured the same original color stimuli. In some cases, however, other objectives for the IDTs may be more appropriate, as discussed in the examples.

ICES from a colorimetric digital camera with nonlinear signal processing and nonideal optical characteristics

In this example the input again is a digital camera that is inherently colorimetric, with responsivities matching those of the CIE 1931 Standard Observer. However, in this case, code values output from the camera are not linearly related to the sensor exposure levels, and the camera has non-ideal optical characteristics.

Nonlinearity is not unusual in actual cameras. Although virtually all electronic sensors are inherently "photon counters" that produce linear analog electronic signals, a camera's analog-to-digital converters may not be perfectly linear in their behavior. In addition, the manufacturer may deliberately choose to introduce a specific nonlinearity in the digital signal processing. The nonlinearity can be subtle or it might be significant if, for example, the camera code values are meant to correspond to HDTV digital video standards. If the camera signal processing is not linear, compensation for the nonlinearity can be included in the input signal processing transform. This can be implemented in an IDT by the use of one-dimensional lookup tables (1-D LUTs). In some cases, the departures from linearity might be somewhat different for each color channel. If so, the respective 1-D LUTs of the color channels would have to be correspondingly different.

Additionally, in most actual digital cameras, an image of a spatially uniform stimulus will not produce equal digital code values for every pixel location. Exposure values will differ due to optical nonuniformities (e.g., lens fall-off) and, if not corrected within the camera, dissimilar sensor responsivities at different pixel locations. The effects of these characteristics can be compensated for by the use of an appropriate 2-D LUT, as shown in Figure 21.4.

The construction of these transforms requires some effort, of course, which raises the question of whether the corrections they impart really are necessary. Certainly when the ICES is defined in terms of linear values, any major camera signal processing nonlinearity (such as a video camera "gamma correction" characteristic) must be accounted for in the input signal processing transform in order to produce ICES image values that are fundamentally compatible with those from other input sources. In less extreme cases, however, the question is more application specific and philosophical.

If the objective were to make the camera serve as a colorimetrically accurate imager for an application such as medical diagnostics, then, yes, every effort should be made to compensate for factors that would reduce the accuracy of the specific device. For theatrical motion picture work, however, that may not be necessary, and it may not even be desirable.

For example, two electronic camera models may be nearly identical and have signal processing that is substantially linear, but each may depart slightly from perfect linearity in different ways. As a result, each camera will produce images having a different "look" or "personality." Use of camera-specific input transforms would eliminate those subtle departures from linearity, effectively making images from the two cameras virtually identical. That generally would be desirable for projects that involve forming composite images from the two cameras. However, eliminating the distinction between the cameras might not be acceptable to a cinematographer who has a preference for one camera's particular "look" over that of another.

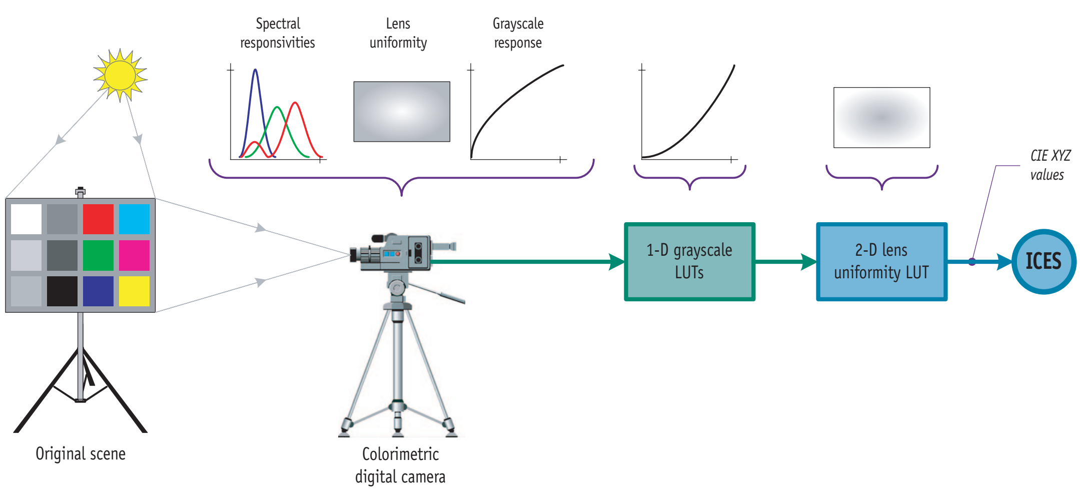


Figure 21.4 Use of 1-D and 2-D signal processing LUTs to compensate for nonlinearities and optical nonuniformities in a digital camera.

Similarly, the vignette effect produced by lens falloff may be aesthetically pleasing in some types of images. The point here is that the question of whether or not to use *all* of the capabilities that can be provided by input signal processing transforms is as much about art as it is about technology. In the transformation to scene space, individual "personality traits" of the input devices and media can be removed to make all inputs essentially identical to the reference device and thus to each other, or they can be left in place—partially or completely—to preserve individual looks. The underlying philosophy of this digital cinema system is that such decisions must be left to the end users and not be imposed by the system itself.

ICES from a colorimetric digital camera with non-CIE responsivities

In the previous examples, the digital cameras being considered were inherently colorimetric devices, i.e., they had spectral responsivities corresponding to those of the CIE 1931 Standard Observer. Because these responsivities are all positive (i.e., they have no imaginary negative lobes), in principle it should be possible to build a camera having such responsivities. That feature certainly would simplify the task of generating CIE *XYZ* values for the ICES. However, to the best of our knowledge, no existing digital motion picture camera has such spectral responsivities. There are engineering reasons, primarily related to the significant overlap of the "red" and "green" responsivities, why building this type of camera is somewhat problematic.

However, practical colorimetric cameras can be built using spectral responsivities corresponding to some *other* set of CIE color-matching functions having less spectral overlap. This is, in fact, a tactic frequently employed by camera manufacturers. In this input example, then, we will assume that the camera responsivities exactly correspond to a set of CIE color-matching functions, but to a set other than those of the CIE 1931 Standard Colorimetric Observer.

A discussed in Chapter 1, a fundamental principle of color science is that all sets of color-matching functions are simply linear combinations of all other sets. That means any set can be transformed into any other set using an appropriate 3×3 matrix. Moreover, it means that a 3×3 matrix can transform the tristimulus values of one set of color-matching functions to visually matching tristimulus values of another set, as explained in Chapter 5. All that is required, then, to allow the camera of this example to be used for input to the system is the inclusion of an appropriate 3×3 matrix in the input signal processing transform. The matrix must operate on image values expressed in linear exposure space, so it must be implemented following any required linearization transformations.

Determining the coefficients of this 3×3 matrix is a textbook procedure. A given set of color-matching functions will be associated with a particular set of RGB primaries for which the chromaticities can be specified. Derivation of a transformation matrix from *RGB* values to CIE *XYZ* values is then a simple algebraic exercise (see Appendix H).

ICES from a noncolorimetric digital camera

The digital camera of this example, unlike those in the preceding examples, is not inherently colorimetric. This means its spectral responsivities do not correspond to the color-matching functions of the CIE 1931 Standard Colorimetric Observer, nor do they correspond to any other set of visual color-matching functions.

This example is entirely realistic because camera spectral responsivities always differ, at least to some extent, from actual color-matching functions. There is a long list of reasons why manufacturers make deliberate departures from color-matching functions. (It is *not*, as some have suggested, due to a lack of understanding of the color science involved.) Regardless of the reasons, the result is that for virtually all real cameras, the encoding problem becomes one of deriving colorimetric information from a device that is not inherently colorimetric. Doing so is, of course, theoretically impossible. Nevertheless, in practice it has to be done anyway.

Before going any further, it is important to emphasize that this complication is not caused by the use of a scene-based ICES. Using a different type of CES, even one based on rendered output colorimetry, would not alter the fact that there are fundamental image-capture color problems caused by the camera spectral responsivities. A scene-based ICES just makes those camera problems more explicit and more directly quantifiable in terms of standard colorimetric measures.

How can *RGB* values from a noncolorimetric camera be transformed into colorimetric values? The process of creating an appropriate transform, generally in the form of a 3×3 matrix, is more of an art than a science. For example, it might seem reasonable to derive a matrix "scientifically" by performing a regression between camera linear *RGB* values and corresponding CIE *XYZ* values for a set of test stimuli. A linear regression relating the two sets of values would generate a matrix that minimizes the mathematical *XYZ* errors; however, it is unlikely that the matrix would produce the most pleasing images. In practice, camera manufacturers have proprietary mathematical tools, image simulation capabilities, and experienced people who can determine matrices or other transformations that minimize undesirable color effects. Optimizing such transformations inevitably involves trade-offs. For example, important memory colors such as neutrals, skin tones, foliage, sky, etc., generally are given greater consideration than less critical colors.

ICES from photographic films

The examples discussed to this point have been for digital cameras. Digital cameras directly provide image signal values in digital form, and their relationships between digital signal values and exposure values generally can be characterized without much difficulty. Therefore it is relatively straightforward to generate ICES values for these devices.

The situation is quite different when photographic films are used for input. First, deriving digital image values from film requires the additional step of scanning. Moreover, film images are formed by chemical signal processing mechanisms that often are quite complex. As a result, determining transforms relating scanned film values to corresponding exposure values can be difficult. In the sections that follow, we will discuss methodologies that address these issues and describe how scanned film values can be transformed successfully to film exposure values and to ICES colorimetric values.

ICES from black and white films

The process used to derive ICES values from black and white films is a good place to begin this discussion because the relationship of image values and exposure values is relatively straightforward for these films. The first step of the process is, of course, scanning with an appropriate film scanner to determine the optical density of each image pixel. This is quite simple for black and white films because their images are composed of silver, which is nearly spectrally nonselective. That makes densitometry straightforward because density readings are essentially independent of what wavelengths are used for measurement. Like any densitometer, however, a film scanner requires periodic calibration (e.g., adjustments of electronic gains and offsets). Such calibration, which often is automatic, is standard on virtually all scanners. Some scanners also provide correction for field nonuniformity. That correction can be important for very critical work, but it generally is not essential for most imaging applications.

The relationship of (calibrated) scanned density values and corresponding exposure values for a particular black and white film can be characterized by scanning an image of a grayscale target of known exposure values. The results can be used to construct a characteristic curve relating scanned density and relative log exposure, such as that shown in Figure 21.5. This characteristic curve is the basis for the 1-D LUT input signal processing transform required for encoding.

The dashed line in Figure 21.5 illustrates an example transformation of a scanned density value to a corresponding relative log exposure value. That simple transformation is essentially all that is involved in the scene-based input-encoding method for black and white films. As discussed in Chapter 12, the transformation is most accurate if it is specific to the particular film being scanned. Alternatively, the individual "personalities" of various input films can be retained if images are processed through a "universal" transform representative of the basic type of film being used. For example, a film having higher-than-average photographic gamma still will produce

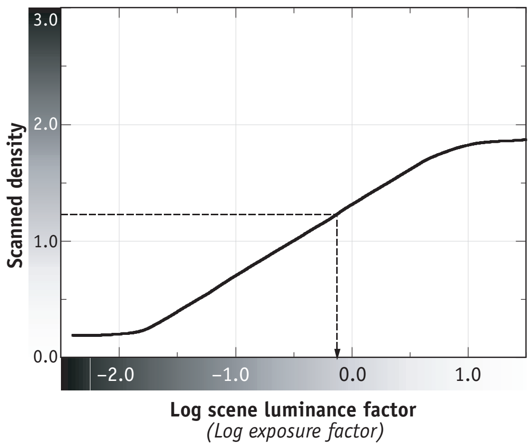


Figure 21.5 A characteristic curve relating scanned density and relative log exposure for a representative black and white film. The dashed line illustrates an example transformation of a scanned density value to relative log exposure value.

reference ICES values if it is input through a product-specific IDT. If instead it is input through a universal IDT based on the photographic gamma of an average film, the film's higher gamma will be reflected in the ICES values and ultimately will result in higher gamma output images.

The remaining steps of the encoding involve only data metric conversions. This might include a normalization to ensure that a properly exposed scene reference stimulus (usually a reference neutral) will have a correct log exposure value, an exponentiation to convert from logarithmic to linear exposure values, and finally a conversion to ICES data metric values. If the ICES were encoded in terms of *XYZ* values, the *Y* tristimulus values would be set equal to the film's normalized exposure values and the corresponding *X* and *Z* tristimulus values would be computed so as to produce the chromaticity specified for the ICES reference white.

ICES from color-negative films

In all of the examples discussed to this point, the digital cameras and black and white films have inherently had a characteristic we refer to as *channel independence*. "Independence" means that an imaging channel (e.g., red) produces a signal that is detectable and separate from the signals produced by other imaging channels (green and blue).

Black and white films are channel independent by definition because they have only one imaging channel. They capture a single channel of exposure information and produce a single measurable signal in the form of a silver image. A monochrome electronic camera is similarly channel independent. Color electronic cameras also are inherently channel-independent devices because each of the red, green, and blue image-capture channels produces its own detectable electronic signal at the output of the image sensor(s) without affecting the signals produced by the other channels. That independence simplifies the process of transforming image signal values to exposure values. (Note, however, that this channel independence is lost, or at least masked, if a matrix or some other form of 3-D signal processing mixes the captured image signals before they are output by the camera.)

If the imaging channels of color photographic films were likewise channel independent, the process of determining scene exposure values from scanned density values would be essentially the same as that for black and white films. The only difference would be that for each image pixel, there would be three of everything: three (*RGB*) density values and three corresponding *RGB* exposure values that would be determined by using three respective *RGB* characteristic curves that make up the film grayscale. That process is illustrated in Figure 21.6.

The illustration indicates that red exposure values can be determined from red density values, green exposure values from green density values, and blue exposure values from blue density values. However, that would be valid only if the measured density values for the three color channels were completely independent. Unfortunately, such independence is not inherent in color photographic films. Therefore measured red density values, for example, while predominantly indicative of red exposures, also are related to some extent to exposures that may have been captured by the green and/or blue imaging layers of the film. For the mapping process shown in the figure to be valid, any density channel interdependence present in both the scanned image data and the characterization grayscale curves first must be

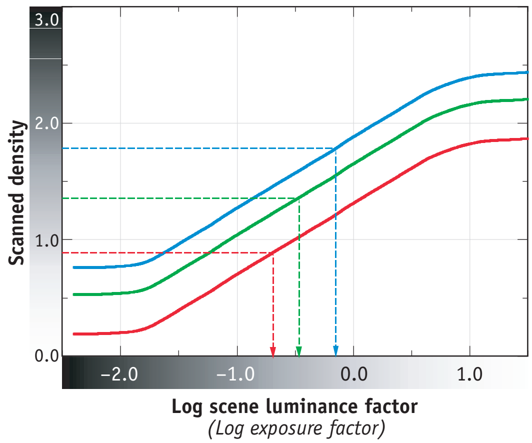


Figure 21.6 If measured density values for the three channels of a color film were independent, exposure values for each channel could be determined directly from respective density values.

characterized and corrected. To understand how that can be done, we need to look at how films capture and record color.

There is just one stage in the process of forming a photographic color-film image when the color-channel signals are independent, and that is when the film has been exposed but not yet chemically processed. At that stage the exposure recorded in the red-sensitive layer is, by definition, the red exposure. That red exposure signal is exactly what a scene-based encoding process is intended to determine. Similarly, the exposure signals for the green and blue channels are recorded in their respective imaging layers. These channel-independent exposure signals are recorded in the form of a latent image that is undetectable by conventional measurement techniques. Chemical processing therefore is required to amplify these microscopic exposure signals and form measurable color images. It is in that chemical process that color-channel interactions first occur. Thus, for example, the amount of cyan dye formed in the film image is due not only to the amount of red-layer exposure, but also to the amount of dye formed in each of the other two layers due to their respective exposures.

Subsequently, the film image must be optically measured for input to the system. That process creates additional color-channel interactions, even in the absence of any prior chemical interactions. Assume, for a moment, that the film and chemical process were designed such that no chemical interactions were produced. Exposure in the red-sensitive layer of the film then would result in the formation of cyan dye, and no magenta or yellow dye would be formed. Thus the cyan dye image itself would be independent of the other channels. However, scans or other optical measurement of that dye would create color-channel crosstalk in the measurements themselves because, in addition to having density to red light, the cyan dye most likely would have some density to green light and to blue light.

The problem, then, is that the use of grayscale-characterization curves alone for transforming color-density values to corresponding color-exposure values requires the density-to-exposure relationships to be 1-D (i.e., channel independent). As just discussed, however, chemical and optical interactions among the color channels instead cause those relationships to be 3-D. To solve this problem, a 3-D transform (e.g., a matrix, multidimensional polynomial, or 3-D table) can be used to computationally remove the *net* effect of the chemical and optical crosstalk interactions from scanner measurements of grayscale-characterization images. This generates a set of channel-independent grayscale-characterization curves appropriate for use in the film input signal processing transform. Applying the same 3-D transform to all scanned image pixel values likewise transforms them to channel-independent density values that can be used with the channel-independent 1-D grayscale-characterization curves to determine (channel-independent) film exposure values. That process is illustrated in Figure 21.7.

To complete the encoding process to an ICES based on scene-space CIE colorimetry, computed film exposure values must be transformed to standard colorimetric values. Because the spectral sensitivities of photographic films, by design, do not correspond to any set of visual color-matching functions, the transformation of film exposure values to CIE colorimetric values is not straightforward. An equivalent problem was described earlier in the discussion of noncolorimetric digital

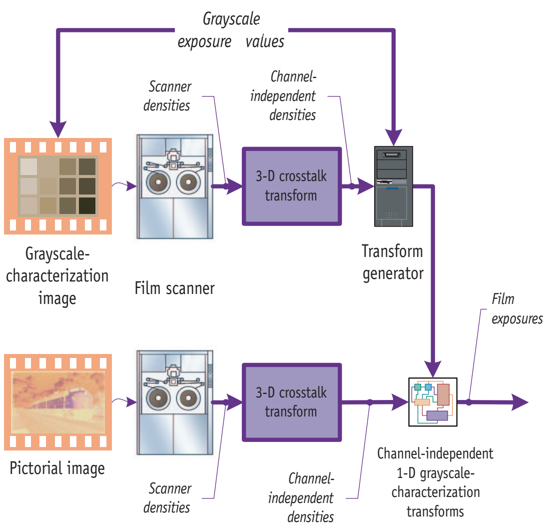


Figure 21.7 Use of a 3-D transform to generate a set of channel-independent grayscale-characterization curves appropriate for use in a color-film input transform. The same 3-D transform is applied to scanned image values to transform them to channel-independent density values that are then used with the channel-independent grayscales to determine film exposure values.

cameras. Transform optimization techniques similar to those used for such cameras are also used in the development of exposure transforms for color photographic films.

As discussed earlier in this chapter and in Chapter 12, input signal processing transforms to scene space can be product specific or universal. In a digital cinema application, the use of product-specific transforms for color negatives would be preferred when the objective is to eliminate the differences among input film stocks. This facilitates a seamless compositing of images from different films. Universal transforms would be preferred when the objective is to retain the "personality" differences among the various films. Since both objectives are important, a digital cinema system must allow and support the use of both types of transforms.

The input signal processing for an image does not necessarily have to stop with its first encoding to ICES, thus decisions regarding universal versus product-specific transforms are not necessarily final. For example, product-specific transforms could be used for all inputs in order to remove their individual characteristics. The resulting ICES values then could be further transformed to restore the characteristics of a particular type of input. This approach allows the retention of those characteristics while also producing a consistent look among all of the inputs.

Finally, it is important to recognize that while various types of scanned densitometric data can be used in developing and implementing product-specific ICES input transforms for color-negative films, input to a universal color-negative film transform must be in terms of printing-density values. That is because the film-to-film differences intended to be retained by the use of a universal transform are represented in a meaningful way only when expressed in terms of printing-density values.

ICES from DPX image files

The principal digital image interchange standards currently in use in the motion picture industry are the Cineon DPX and ANSI/SMPTE DPX file formats. The standards are similar, and both are based on negative-film printing-density values. Therefore DPX files can be input to the example system and transformed to ICES values by the same types of universal or product-specific transforms used with negative-film scanners.

ICES from print films and from other output media

As discussed earlier, the principal forms of input to this example digital cinema system are digital cameras, negative photographic films, computer-generated imagery, and DPX files. However, it is quite likely that on occasion there would be a need to input images from other media, including media that normally function as *outputs*. In particular, image restoration work may require the input of images that only exist in the form of motion picture projection prints. The challenge of meeting that requirement provides an excellent opportunity to demonstrate the flexibility and inherent inclusiveness of systems designed according to the principles of the Unified Paradigm.

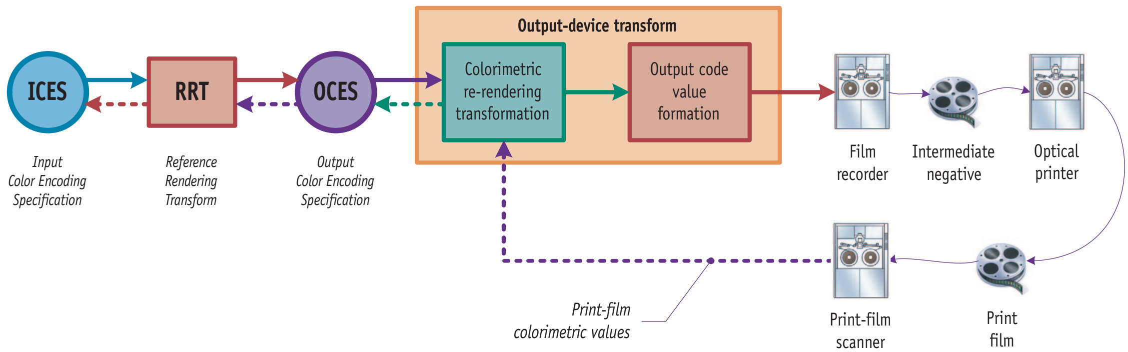


Figure 21.8a A conceptual closed-loop process starting from ICES values, which are used by the system to produce a motion picture print. The colorimetry of that print is then measured, and the measured colorimetric values are transformed such that the starting ICES values are regenerated.

Figure 21.8a illustrates how the system normally would *generate* an image on a motion picture print film. The solid arrows indicate that input (ICES) values would be transformed to reference output (OCES) values and passed to an output-device transform (ODT) for a negative-film recorder. The film recorder would generate an intermediate negative that then would be optically printed to form the print-film image.

As Figure 21.8a also shows, two basic functions (described later in more detail) take place within the ODT. In the first, OCES colorimetric values are re-rendered to colorimetric values appropriate for the print film. The second ODT function determines the film-writer digital code values required to produce an intermediate negative that in turn will produce a print having the specified colorimetry.

Imagine now that the colorimetry of that print film were measured and the colorimetric values were fed back into the system at the location shown in Figure 21.8a. If these values were then processed through the signal processing chain indicated by the *dashed* arrows, they would generate the ICES values from which the print-film image originated. In other words, a closed-loop path would have been followed from ICES values to print-film colorimetry back to the same ICES values.

Now consider that just as the described process can determine ICES values for a print-film image generated by the system, it also can determine ICES values for other print-film images produced by some other means. Thus, colorimetric values measured from any (output) image on motion picture print film can be transformed to (input) ICES values using the path indicated by the dashed arrows in Figure 21.8b. Note that if the print-film scanner is a true colorimetric scanner, as described in Chapter 11, the particular dye set of the print film is not an issue. If instead the scanner is a densitometric scanner, then appropriate transformations would be required to determine colorimetric values from scanned densitometric values (also discussed in Chapter 11).

This raises an interesting question: do ICES values determined by this print-input process represent the original scene initially photographed to make that print? There is no way to know for certain whether they do or not without having access to complete information regarding the print film, the optical printer, the intermediate film, the film writer, all of the transforms involved, the original camera film, any image adjustments made, etc. We would submit, however, that it also does not matter.

The described process will have generated a digital image, in terms of ICES values, based on this particular digital cinema system, with its known image processing paths, and with a known output medium. That means two things. First, if the ICES image is processed normally through the system, it will generate a new image that is a visual match to the original print-film image. That in itself is a significant

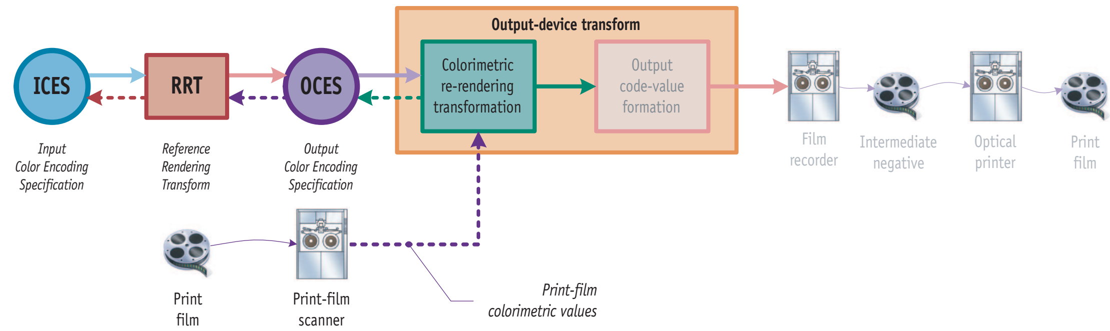


Figure 21.8b Practical use of the methodology of Figure 21.8a to input a motion picture print to a digital cinema system. The print is scanned to determine image colorimetric values that are then processed to form ICES values (dashed arrows).

capability. Second, and perhaps more importantly, derived ICES values represent scene-space colorimetry that currently would generate the same appearance as that print-film image. As such, those ICES image values are entirely compatible and interchangeable with all other ICES image values derived from more conventional inputs. Therefore ICES images generated by this process can be seamlessly intercut with other ICES images, they can be modified using the same image processing tools, and they can be processed by the system for output to any type of display device or medium included in the system.

Images from motion picture print films were chosen for this example because it seems reasonable that they would be the most likely type of output images used for input. However, the same basic procedure can be applied to any hardcopy or softcopy output image for which colorimetry can be determined, and it also can be used to create ICES values from data files of output-image colorimetry. The procedure is particularly straightforward to implement for images from output devices or on media comparable to those already included in the system. In such cases, the required signal processing path from output colorimetry to ICES can be created by inverting the respective ODT and the RRT.

ICES from computer-generated images

Computer-generated images can be incorporated into the system in a number of ways and at several points within the system. Preferably, they would be brought into the ICES in a way that ensures they are compatible and can be intercut with images input from all other sources. The process for doing that is described below.

Computer-generated images could, in principle, be created and manipulated directly in terms of ICES values. It is more likely, however, that practitioners using various CGI techniques will prefer to work in terms of other color spaces that are familiar and perhaps more intuitive or otherwise better suited for the particular tasks being performed. For that reason, Figure 21.9 includes a transform relating the

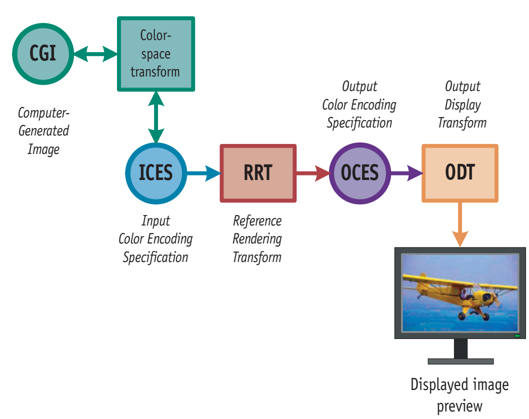


Figure 21.9 Use of a reversible transform to convert between ICES values and a user-specified CGI space. The display provides a real-time preview of final image appearance.

ICES and a user-specified CGI color space. The two-way arrows indicate that images may originate in the CGI space and then be transformed to the ICES and also that ICES images can be transformed to the CGI space for inclusion in work being done there and then returned to the ICES.

Because the ICES directly specifies input-image scene-space colorimetry, transformations to and from commonly used CGI spaces are straightforward. If we again assume, for example, that ICES values are expressed in terms of linear CIE *XYZ* tristimulus values, then transformations to various linear RGB and YCC (luminance/chrominance) spaces commonly used for CGI require only 3×3 matrices. Transformation of ICES values to video code values, which sometimes are used in CGI work, would only require a matrix and a 1-D LUT. Such transformations are simple and fast, and when properly implemented they are reversible. Since ICES values are assumed to contain a normal amount of camera flare, CGI images can be made more compatible with actual camera images if a similar amount of camera flare is computationally added during the CGI creation process. The colorimetric effects of camera flare can be computed using operations similar to those used to determine the effects of viewing flare (Appendix E).

Also included in Figure 21.9 is a display device with its associated output signal processing from the ICES. The display provides a real-time preview of final image appearance. Monitoring this appearance is of great value in image creation and helps ensure the compatibility of CGI-created ICES images with ICES images from other forms of input.

ICES reference viewing conditions

The input signal processing transformations described so far were simplified by stipulating that the original-image-taking conditions were identical to the ICES reference viewing conditions. This provision avoided appearance-related issues and allowed us to focus on other aspects of the input-encoding process. In practice, of course, differences between actual image-taking conditions and ICES reference viewing conditions will be common and must be addressed.

To be consistent with the concept of scene-space encoding and customary motion picture practices, the specifications for the ICES reference viewing environment of this system should be representative of those of live scenes viewed in average daylight. Accordingly, viewing flare should be specified as zero because any flare light present in a scene would be an integral part of that scene and will be included in its colorimetric description. As discussed earlier, however, ICES images are assumed to include a degree of camera flare representative of motion picture cameras. Compensation for that camera flare is incorporated in subsequent image processing stages of the system and will be discussed later. The ICES surround type should be specified as normal or average because, in a typical scene, objects ordinarily are surrounded by other similarly illuminated objects. Finally, the luminance and chromaticity of the observer adaptive white should be specified to be consistent with average daylight conditions.

As briefly discussed earlier, ICES specifications do not dictate a single set of conditions under which all original scenes must be captured; realistically, actual image-capture conditions often will differ from the ICES reference viewing conditions. In those situations, input-image colorimetric values must be transformed *appropriately* to account for the effects that such differences will have on the specified image colorimetry and on an observer's perception of that colorimetry.

The word "appropriately" is emphasized in the preceding sentence because there is much more involved here than strict colorimetric matching or even color-appearance matching. In fact in some situations where the image-taking conditions and ICES viewing conditions differ significantly, transforming "appropriately" actually may mean "doing nothing" to captured scene colorimetric values. Although that may seem inconsistent with much of what has been said previously regarding color-appearance encoding, it is not.

Consider, for example, a situation where a camera has flare greater than the amount assumed in the ICES. In some cases this may be judged as a detriment, and thus it would be appropriate for the camera's input signal processing transform to include compensation for the additional flare. In other cases, however, a particular camera or lens may be selected expressly *because* its flare effects (e.g., reduced luminance contrast and lower color saturation) are considered aesthetically desirable for a particular scene. In that case, it would be completely inappropriate for the system to automatically compensate for the additional flare. If simply left uncorrected in the input signal processing transform, the camera's additional flare will be represented appropriately in terms of ICES values, and the desired effects of that flare ultimately will appear in the output images.

As another example, consider a scene that has been photographed under tungsten illumination. Because the ICES reference viewing environment includes an observer adaptive white having a chromaticity corresponding to that of daylight, it might be assumed that some form of color-balance correction must be made in order to compensate for the use of the "wrong" taking illuminant. For example, a white-balance adjustment could be made in a digital camera, a color-compensating filter could be used over the camera lens, a tungsten-balanced film could be used instead of a daylight-balanced film, or a chromatic adaptation transformation could be included as part of the input signal processing. However, the assumption that some form of compensation is required may or may not be correct.

If the tungsten illumination had been used on a stage set meant to simulate an outdoor scene, then certainly some form of correction would be required in order to make the tungsten-illuminated images match those from actual daylight-illuminated scenes in the ICES. In a different situation, however, the tungsten illumination may be considered part of the scene itself. For example, the scene might be a living room or other setting where tungsten illumination normally would be expected. Accordingly, the objective would no longer be to match the color appearance of outdoor scenes. Instead, the goal would be to create ICES images that retain an appropriate amount of the warm color balance associated with tungsten illumination. Similar arguments can be made for other situations such as night scenes, sunsets, simulated fantasy worlds, etc., where the objectives of color-appearance encoding are subject to various creative interpretations.

Perhaps the most dramatic illustration of why scene-based input color encoding can involve more than strict color-appearance matching when different viewing conditions are involved is in what is known as *day-for-night* photography. In this technique, scenes are captured in daylight but are made to appear as if they had been captured at night. An impression of darkness is created by the use of filters, deliberate underexposure in the camera, and other photographic tricks. There in fact are *three* viewing environments involved here: the actual image-taking environment (broad daylight), the simulated environment (nighttime), and the ICES reference environment. Clearly in this situation, it would be completely inappropriate for the input signal processing to attempt to "fix" the day-for-night shots to make them appear as they would in either the actual taking or the ICES environment. An appropriate input signal processing transform instead would do nothing to alter the color appearance deliberately created by the cinematographer.

The most important point of this discussion is that the only purpose of the ICES viewing environment specification is to provide a means for unambiguously specifying the appearance of colorimetric values *in the ICES itself*, at that particular stage of the imaging system. *The ICES specifications should not be misinterpreted as a set of environmental conditions under which images must be captured, nor should they be misinterpreted as a fixed set of rules as to how images must be processed for input.*

Decisions regarding actual image-capture conditions and how those conditions should influence the transformation to ICES are part of the creative process. The underlying philosophy of this system is that all such decisions must be left to the discretion of the system's users and must not be imposed by the system itself.

ICES connections to other workspaces

The final consideration related to the encoding method of the ICES is its link to other color-image workspaces where various types of image manipulations can be performed. Depending on the data metric chosen for the ICES, some types of image manipulations might be performed directly in the ICES color space. However, as in the case of computer image generation, practitioners likely would prefer to

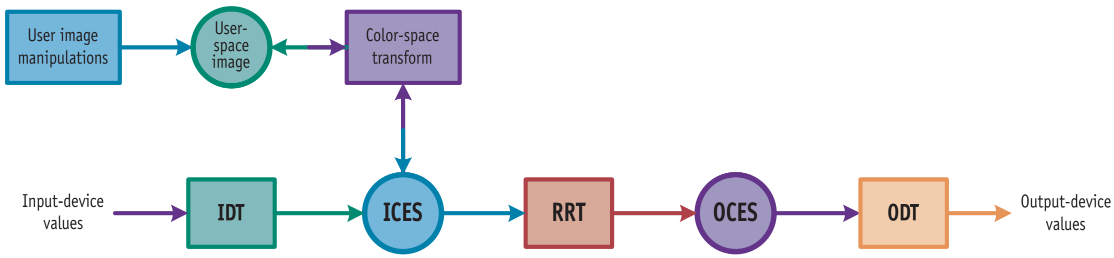


Figure 21.10 A data path including a color-space transformation from the ICES to another color space where user-specified manipulations operate on an image. Upon completion, a modified image is processed through an inverse transform to return to ICES space.

work in various existing color spaces that are familiar and best suited for particular types of imaging operations. Spaces based on perceptual attributes such as lightness, hue, chroma, and saturation, for example, are useful for adjusting colors of individual objects. Other spaces are better suited for merging portions of multiple images into seamless composite images and other image manipulation tasks.

Figure 21.10 illustrates a data path that includes the use of color-space transforms from the ICES to another color space. User-specified manipulations then would operate on the image in that alternate space. Upon completion, a modified image would be processed through an inverse transform to return to ICES space. From there, the modified image would continue as normal through the system.

Some types of image adjustments, however, are best applied before input-image values have been fully transformed to ICES values. In particular, it often is important to have direct access to input-device/medium *RGB* exposure values prior to their conversion to CIE colorimetric values. A color-management application implementing this system preferably would provide that access. There are several reasons for this preference.

First, it is very likely there will be errors in the derived *RGB* exposure values. Errors can result from under- or overexposure in the original photography, color-balance errors due to imperfect white-point adjustment of an electronic camera or illuminant/film color-temperature mismatch, differences among the devices and/or media that are not fully accounted for by input calibration, and possible errors in the transformation process itself. Because these errors occur in the exposure space of the particular input device or medium, they are most easily corrected in that same space rather than later on in the ICES space. How much more difficult they would be to correct in the ICES will depend on its data metric.

Assume, for example, that due to an incorrect white-point adjustment, the overall red exposure of a digital camera image is high relative to the green and blue exposures. This is easy to identify and correct in the camera's own RGB exposure space. All that is required is a rescaling (downward, in this example) of all red linear exposure values. If the uncorrected image is instead first transformed to an ICES based on another linear RGB space or to CIE *XYZ* space, the high red exposure values would produce changes in all three ICES color channels as a result of the 3×3 matrix involved in the transformation. So what began as a simple one-channel color-balance problem would become a more complex three-channel problem. The ICES image still may not be too difficult to color-balance, but some rescaling of all three channels would be required.

Considering this example further, if the uncorrected image were first transformed to an ICES based on a nonlinear space such as CIELAB, the high red exposure values would affect all L^*, a^*, and b^* values. If the $L^*a^*b^*$ values for the color-balanced and unbalanced images were plotted against each other, the impression would be that something quite complex has occurred. Most likely, a correctly color-balanced image could not be produced simply by applying overall scales or shifts to the unbalanced image's $L^*a^*b^*$ values. Adjustments that perfectly correct some colors may not correct others. Again, what started out as

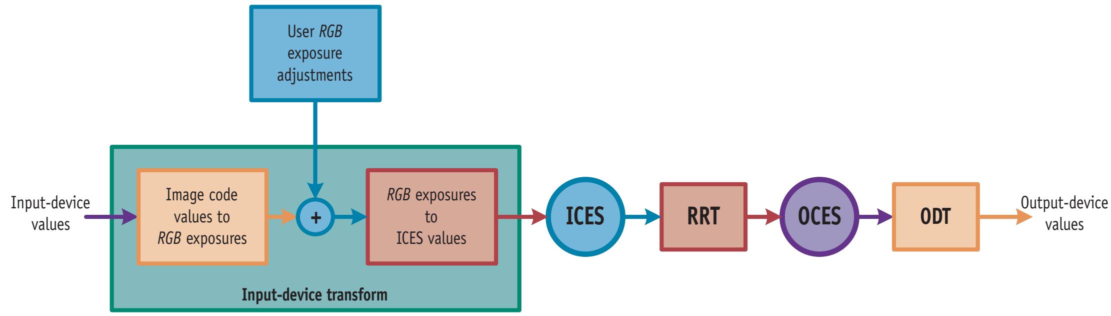


Figure 21.11 Use of a two-part input-device transform (IDT) provides direct access to *RGB* exposure values for adjustment.

a problem that could have been corrected by a single 1-D adjustment will have become a complex 3-D problem that can be quite difficult to resolve.

Our experience is that it generally is best to fix problems in the same space in which they occur. In this case, that means addressing device and media exposure-space errors prior to an input image's transformation to ICES. In Figure 21.11, access to the input-image *RGB* exposure values is provided by implementing each IDT in two parts. A user (or algorithm) could then manipulate images at the exposure level, as shown in the figure. For these types of input-image corrections, and for other image manipulations as well, it is useful if a real-time preview image can be observed as changes are made.

Rendering for output

In order to best meet the objectives of digital cinema, ICES images are represented in terms of scene-space colorimetric values. As discussed in earlier chapters, when displayed and viewed directly, colorimetrically accurate scene-space images are perceived as appearing "flat" and "lifeless." Image *output rendering* therefore is required to transform scene-space colorimetric values to output colorimetric values appropriate for display. The output-rendering process consists of a package of adjustments that are applied to scene-space colorimetric values. Collectively, these adjustments result in an *image-state transition* from a defined scene space to a defined rendered space.

It is important to recognize that this transition is intended to produce images that are excellent *reproductions* of scenes, not *re-creations* (objective appearance matches) of those scenes. The assessment of image reproductions is influenced by many factors other than those associated strictly with appearance matching, including aesthetic expectations that have developed throughout human cultural history. Centuries of visual art, more than a century of traditional photography, and decades of electronic imaging have contributed to accepted conventions for image reproductions. For example, certain colorimetric modifications of luminance factors and chroma levels in a reproduction are used to *suggest* and *represent* (but certainly not *duplicate*) a brightly illuminated outdoor scene. This is accomplished by relatively subtle colorimetric shifts, not by color matching.

The ultimate intent of the output-rendering process, then, is to produce displayed images that are judged to be of high quality and consistent with the influences, expectations, and conventions that contribute to the interpretation and assessment of all forms of image reproductions. In addition to the artistic aspects of this process, there are many technical factors involved. In fact, if one were to make a list of all of the issues that might be considered in determining an optimum relationship of reproduced colors to scene-space colors, the task of developing an output-rendering transformation might seem overwhelming, if not impossible. For example, there are dozens of known psychophysical effects that influence the perception of color. If all were considered, the process of output rendering would become extremely complex.

Our experience, however, is that under the conditions relevant to practical imaging applications, many psychophysical effects are not very significant and can safely be ignored. Thus only the manageably small number of factors described below needs to be considered. As expected, many of these factors are the same as those previously specified for appearance-based color encoding.

- ***Camera flare:*** Flare light in a still or motion picture camera physically lowers the exposure contrast of the recorded image, especially in shadow areas, and it also desaturates recorded colors. To compensate, rendering can include appropriate adjustments of the image grayscale and chroma levels. Although this correction could instead be applied in every input signal processing transform to ICES, it is somewhat simpler to apply a single representative compensation in a reference output-rendering transformation. The decision as to whether or not to include other compensation in the input signal processing transform for a camera having an atypical amount of flare is then left to the user.
- ***Viewing flare:*** Flare light in the viewing environment physically lowers the luminance contrast of the display, especially in shadow areas, and it also desaturates reproduced colors. To compensate, output rendering again must include appropriate adjustments of the image grayscale and chroma levels.
- ***Image luminance:*** In nearly all situations, the (absolute) luminance of the displayed image will be considerably lower than that of the original. This lowers the perceived luminance contrast and colorfulness of the display. To compensate, output rendering once again must include adjustments of the image grayscale and chroma levels.
- ***Observer chromatic adaptation:*** The perception of color is strongly affected by the observer's state of chromatic adaptation. Output rendering must modify the chromaticities of scene-space colorimetry to be consistent with the observer's state of chromatic adaptation in the display viewing environment.
- ***Lateral-brightness adaptation:*** The perception of image luminance contrast is affected by the relative luminance of the areas surrounding a displayed image. This is of particular importance in motion picture applications because the dark surround of cinema theater projection significantly lowers perceived image luminance contrast. To compensate, output rendering must include appropriate adjustments of the image grayscale.
- ***General-brightness adaptation:*** The perception of image brightness is also affected by the relative luminances of areas surrounding the displayed image. Again, this is of particular importance in motion picture applications due to the darkened theater viewing environment. A well-designed output-rendering process would take advantage of this phenomenon by altering the grayscale characteristic to effectively increase the luminance contrast and dynamic range associated with image highlight information.
- ***Local-brightness adaptation:*** In live scenes, an observer can sequentially focus on and locally adapt to various regions within a scene. This allows details to be seen sequentially in areas of deep shadows and areas of bright highlights. Little or no such local adaptation takes place in the viewing of most reproductions. This adaptation effect is emulated in paintings by localized adjustments of tone reproduction, and similar localized adjustments can be made in digital signal processing. To some extent, the effect also can be simulated by appropriate shaping of the highlight and shadow regions of a globally applied output-rendering grayscale.
- ***Color memory and color preference:*** Optimum output rendering should account for the psychological influences of color memory and color preference. Again, the goal is to generate reproductions that are judged according to subjective standards. In many cases, this involves further modifications in color reproduction to produce displays that, although not accurate, conform to observers' recollections and preferences for color.
- ***Output luminance dynamic range and color gamut:*** The final consideration is that the output-rendering process ultimately must specify colors that are within the luminance dynamic range and color-gamut capabilities of the actual display device or medium.

Reference output rendering

As shown in Figure 21.2, renderings for the various outputs of this system are accomplished in two stages. In the first stage described here, ICES scene-space values are transformed through a single RRT to produce idealized reference-rendered OCES colorimetric values. In the second stage, described in the following section, OCES values are further transformed by output-specific ODTs to produce final output-device code values. The same output results could be obtained, of course, by instead using ODTs that work directly from ICES values. However, there are important advantages to defining a reference rendering transformation for the system. In particular, a uniquely defined RRT provides an unambiguous specification of how ICES-encoded colors would appear if they were ideally output rendered and displayed with essentially no restrictions.

It is useful to recall the debates that occurred when the technology for "colorizing" black and white motion pictures was introduced. Controversies quickly arose regarding the concept of original intent. Would color have been used if it had been available at the time? If so, what color "look" would have been chosen? Strong, vibrant colors or muted, subdued colors? A warmer or a colder overall color balance? Those issues could not be resolved because there was no means of knowing for certain the original artistic intent. In the system described here, the RRT provides that means. It allows images to be archived—directly in terms of OCES values, or optionally in terms of ICES values together with the RRT—such that there can be no uncertainty as to their intended output appearance. The RRT, then, provides an important complement to the ICES.

Although the RRT is intended to represent an idealized rendering of ICES images, there are practical issues that must be considered in its design. In particular, the rendering must not be so idealized that it effectively just transfers the entire rendering task to each of the various ODTs. The reference rendering characteristics therefore must be somewhat realistic in order to minimize the degree and complexity of those transforms. However, that requirement must be balanced with an important OCES objective of retaining information that someday might be used by future types of display devices. An additional consideration is that the RRT transform must be accurately reversible in order to provide a means for output images to be brought into the system and transformed to ICES when necessary.

With these considerations in mind, the RRT designed for this example system is based on a hypothetical display device having a luminance dynamic range greater than 10 million to one (a visual density range greater than 7.0). Basing the reference rendering on this highly enhanced grayscale characteristic makes possible the encoding of significant highlight and shadow information in the OCES. This ensures that the OCES can contain more than enough information needed for all existing output devices and media. In addition, it ensures that sufficient output-rendered information will be available in the future if the maximum luminance levels and dynamic ranges of display devices were to increase. For comparable reasons, the RRT is based on a set of color primaries having a color gamut that encompasses the chromaticities of all colors.

Figure 21.12 illustrates the effects of rendering a scene-space grayscale (the line of unity slope in the figure) according to the 1-D processing of the example RRT. Note that the reference rendered grayscale has an overall nonlinearity (in log–log space). This nonlinearity is necessary to compensate for camera flare understood to be included in the ICES data and also to compensate for viewing flare anticipated to be present in motion picture theaters.

The reference rendered grayscale includes an overall slope increase that provides compensation for the perceived reduction in luminance contrast induced by the relatively low luminance of projected images. In addition, the grayscale has a further increase in slope to compensate for lateral-brightness adaptation effects induced by dark-surround projection. The grayscale also adjusts overall output visual density values in a way that is consistent with the visual effects of general-brightness adaptation. As discussed earlier, when applied properly this adjustment yields negative visual density values in the extreme highlight regions of the grayscale (at log scene luminance-factor values greater than zero). These negative visual density values correspond to CIE Y and L^* values greater than 100, which appropriately describe the "above-white" appearance achievable in highlights displayed in a darkened theater.

Finally, the classic *S*-shape of the grayscale results from the deliberate compression of highlight

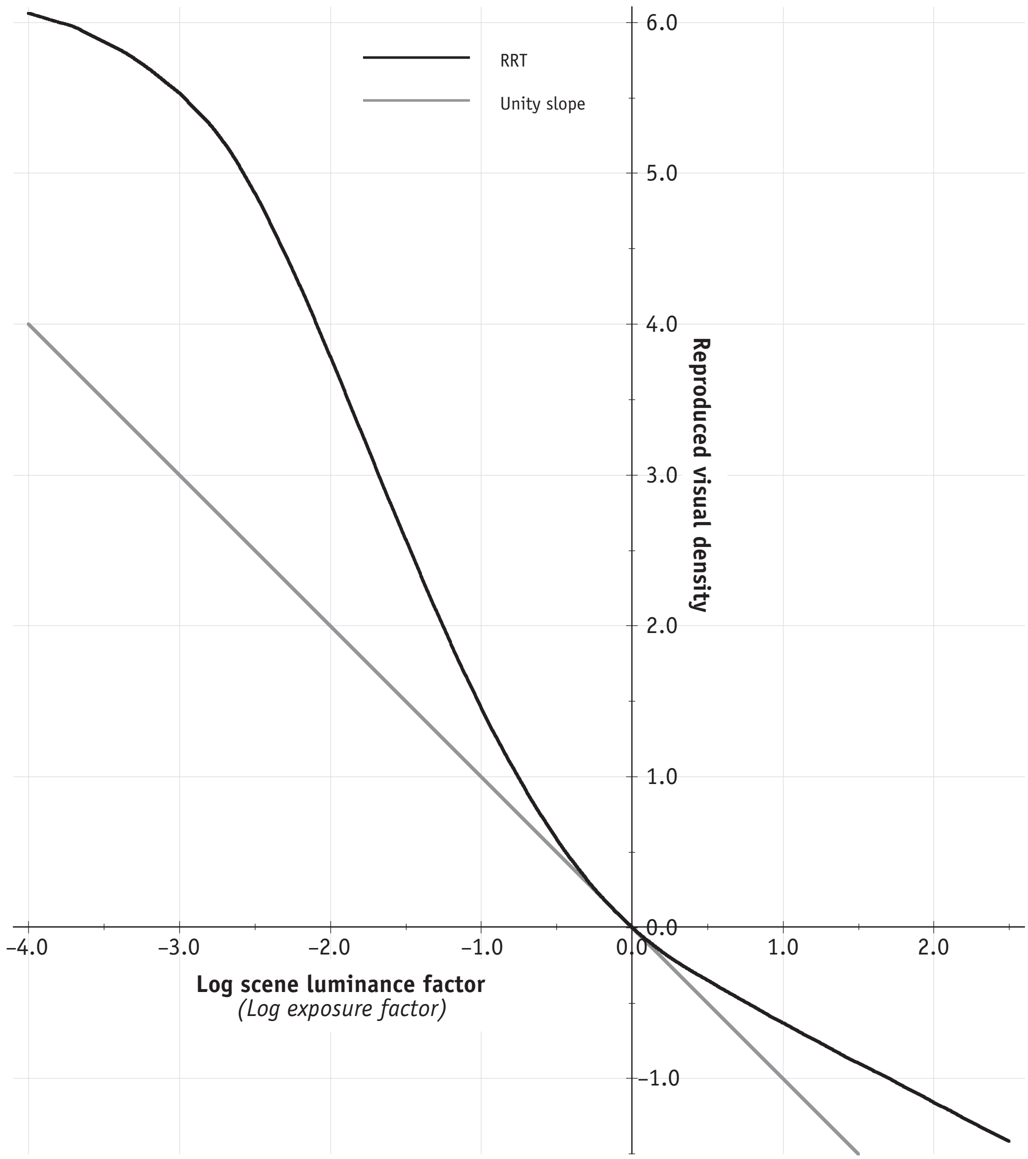


Figure 21.12 An example reference rendering of scene-space grayscale values for an idealized digital cinema projection. Reference rendered values differ significantly from their respective scene-space values for the reasons explained in the text.

information and shadow information in the rendering process, consistent with the luminance dynamic range specified for the reference rendering. This luminance compression avoids any abrupt clipping of information and simulates, to some degree, the effects of local-brightness adaptation.

In this example RRT, the 1-D grayscale transformation is applied to individual RGB color channels derived from the ICES values. Doing so affects how colors, as well as neutrals, are rendered. For the most part, the resulting color-rendering effects are desirable because they tend to compensate for several

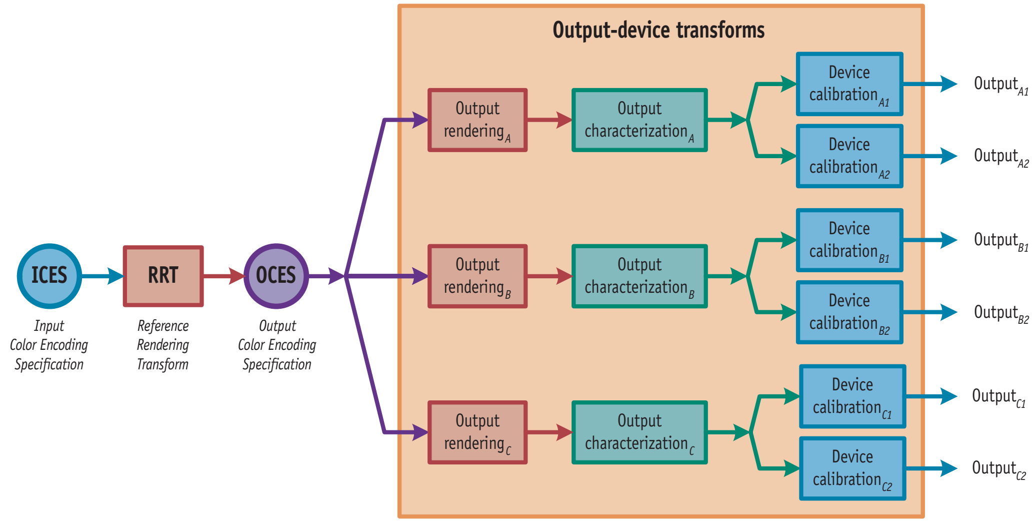


Figure 21.13 In this sequence of transformations, reference rendered OCES images are processed to produce digital code values appropriate for each specific output.

of the physical, psychophysical, and psychological color-related factors previously described.

The exact effect that the transformation will have on colors is determined in large part by what RGB color primaries are used to encode colors at the input to the rendering transform. In general, the transformation will tend to raise the chroma of most colors somewhat too much. The reason is that the transformation is based on an optimum grayscale rendering. As such, it includes compensation (a slope increase) for the perceived reduction in luminance contrast induced by the dark surround. Since a dark surround does not induce a corresponding reduction in perceived chroma, using the same 1-D transform for colors will tend to overly increase their chromas. Compensations for this and any other undesirable color effects produced by the grayscale transformation can be provided by additionally incorporating appropriate 3-D processing as part of the reference-rendering transformation of ICES values to OCES values.

Output from OCES

In the second stage of the output-rendering process, reference-rendered OCES images are processed through a series of transformations that ultimately produce digital code values appropriate for each specific output device and each associated output-image viewing environment. The sequence of this process is shown in Figure 21.13 and discussed in detail below.

As Figure 21.13 illustrates, output signal processing takes place within an ODT and begins with a *second* output-rendering transformation (a *re-rendering*). This transformation adjusts reference-rendered OCES colorimetric values to alternative rendered colorimetric values that are physically realizable by a particular type of output. It is important to recognize that the intent here is not to create anything fundamentally new; it is to deliver the color specified in the OCES as faithfully as possible within the capabilities and limitations of the given output. This is a critical issue: for this system to serve as a means for working with, communicating, and archiving color within the motion picture industry, the meaning of color in the ICES and the OCES must be both consistent and unambiguous. This would not be possible if specified OCES colors were subjected to significant after-the-fact “creative” re-renderings within the ODTs.

In cases where the output has an extensive luminance dynamic range and color gamut (e.g., a photographic print film or digital cinema projector) relatively minor compromises generally will be required in re-rendering OCES colorimetry to output colorimetry. In other cases, however, the amount of dynamic range and/or gamut compression required may be considerable, and decisions regarding trade-offs will need to be made. As discussed earlier, the underlying decision is fundamentally between the Type *B* and *C* paradigms. For example, if the emphasis is on matching outputs to each other (Type *B*), a trade-off might be to match the hue and chroma of a particular color at the expense of sacrificing some luminance detail in that color. If instead the emphasis is on overall output-image quality (Type *C*), the trade-off might be to allow some changes in hue and/or chroma in order to better preserve details in that color. It should be emphasized that these distinctions should be relatively subtle and that the ultimate objective is always to attempt to match the appearance of OCES colors.

When the output capabilities are limited, scene-by-scene, frame-by-frame, or even within-frame adjustments may be required in addition to the global adjustments provided by an ODT. This could be accomplished in various ways, including use of a dynamically variable ODT or by direct editing of OCES values. Another alternative would be to make such alterations in the ICES while previewing images from the actual output or from an accurate simulation of that output. The resulting ICES should, of course, then be designated as being specific for the particular output.

Figure 21.14 is an illustration of the relationship between the grayscale of the example RRT and an example output re-rendered grayscale of a digital cinema system that includes a digital projector representative of the current technology. Note that the grayscales are similar over much of their range, but as expected there is significantly more gradient in the highlight and shadow regions of the RRT.

The process of output re-rendering from the OCES generally is straightforward. For neutral colors, it involves only a 1-D transformation relating the grayscale characteristic of the RRT to a grayscale appropriate for the actual system. Figure 21.15 illustrates the development of a grayscale-mapping transform. The arrows in the figure indicate the mapping of two example data points from their values on the RRT grayscale to their corresponding values on the actual output system grayscale. For the mapping to be correct, the respective grayscales must be registered in two respects. First, they must be "speed balanced," i.e., they must be matched properly along the exposure axis. This is done by aligning the curves based on the relative exposures of known reference test patches. In Figure 21.16, the grayscales of four example systems and the RRT have been aligned along the relative log exposure axis, based on knowledge of the aim visual density that each should produce from a normally exposed neutral test patch of 20 % reflectance (corresponding to -0.7 log luminance factor, as indicated by the dotted line in the figure).

Second, the grayscales must be aligned along the visual density axis such that they are matched for perceived overall brightness. A common assumption is that this can be accomplished simply by aligning the curves such that they match at their point of maximum brightness (minimum visual density). As was discussed in previous chapters, such matching is based on the mistaken supposition that the brightest areas within the visual field determine the observer's state of general brightness adaptation. Numerous psychophysical experiments—and decades of practical imaging experience—have proven that supposition to be incorrect.

With some experience, it usually is possible to align grayscales for equivalent brightness without performing extensive or elaborate experiments. In practice, any alignment errors will show up soon enough once images are produced. If images appear to be properly rendered except that those from one type of output are consistently too dark or too light, it is a simple matter to adjust the associated grayscale accordingly. The most straightforward way to isolate this and other grayscale problems from color-related problems is to use monochrome images.

Another issue related to re-rendering for output is color gamut. To function as a true reference, the RRT and its associated OCES must encompass a color gamut that supports all current and anticipated forms of output. It would be expected, then, that some form of color-gamut mapping would be required in transforming from OCES colorimetry to that of any real output device or medium. Applying the grayscale transform just discussed to each of the color scales expressed in terms of ERIMM RGB

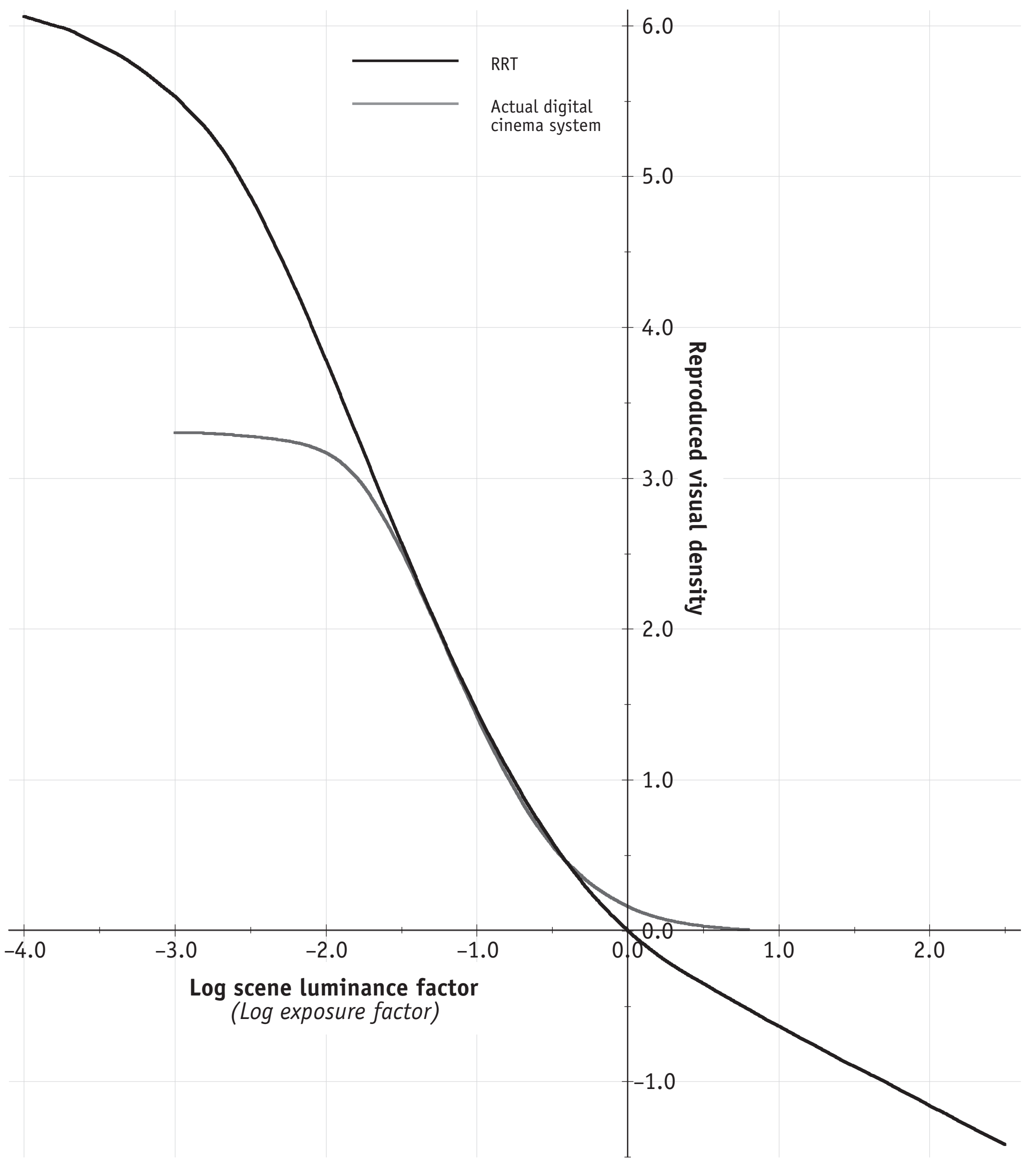


Figure 21.14 A comparison of the grayscales of an appropriate RRT and an actual digital cinema system.

primaries (see Appendix F), or in another space based on similar design criteria, provides some degree of color-gamut mapping. In most cases, however, additional 3-D color-gamut mapping also will be required.

In addition to the grayscale and gamut-mapping operations just described, an ODT must include appropriate colorimetric transformations to account for any differences between the conditions of the OCES reference viewing environment and the conditions of the actual viewing environment associated with the particular output. Although, as discussed earlier, the objective of such appearance transformations is a matter of creative judgment in the case

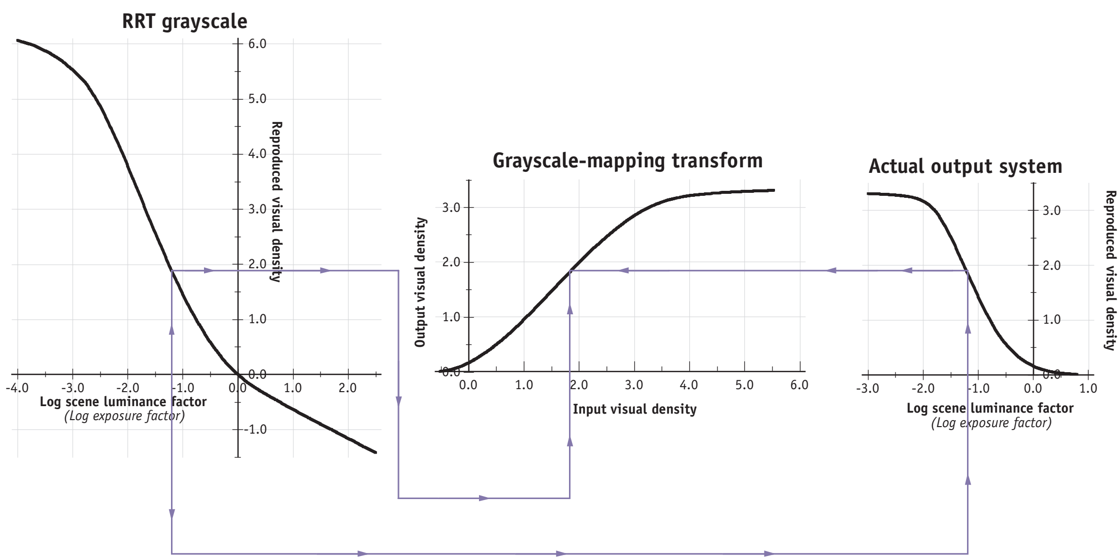


Figure 21.15 Development of a grayscale-mapping transform for output re-rendering of OCES values.

of the ICES, the objective is absolutely straightforward here: it is to consistently display the color appearance specified by the re-rendered OCES values and the OCES reference viewing environment despite any differences in actual output-image viewing environments.

Consistent color appearance across output-image viewing environments can be accomplished by using the colorimetric transformation methods described elsewhere in this book. It also would be appropriate to use color-appearance models here because, unlike rendering from scene space, the objective is straightforward color matching among relatively similar viewing environments.

Output characterization and calibration

To summarize the color signal processing to this point, finalized ICES values were transformed through the RRT to form OCES values. The OCES values then were re-rendered and gamut mapped, as necessary, in each device-specific ODT to form aim colorimetric values for the specified output device/medium and output viewing conditions. A complete system would, of course, include multiple ODTs, each associated with a particular type of output device or medium.

As shown previously in Figure 21.13, the next step in the output signal processing is output characterization. *Output characterization* is a procedure for defining the colorimetric characteristics of a single device or a reference device that is representative of a group of devices used for output on the imaging system. A colorimetric output characterization transform can be developed from device/media models or from empirical data obtained by measuring the colorimetric values of an appropriate number of color patches produced on a reference output device from an array of output-device code values (Figure 21.17).

An *output calibration* transform corrects any deviation of a particular output device from the reference device on which the characterization transform was based. This combined characterization/calibration approach has a number of advantages for digital cinema. In particular, since the bulk of the transformation is performed in the characterization transform, calibration transforms generally can be quite simple. In most cases they can be derived and updated using a relatively small set of test colors, which makes the calibration procedure fast and inexpensive to

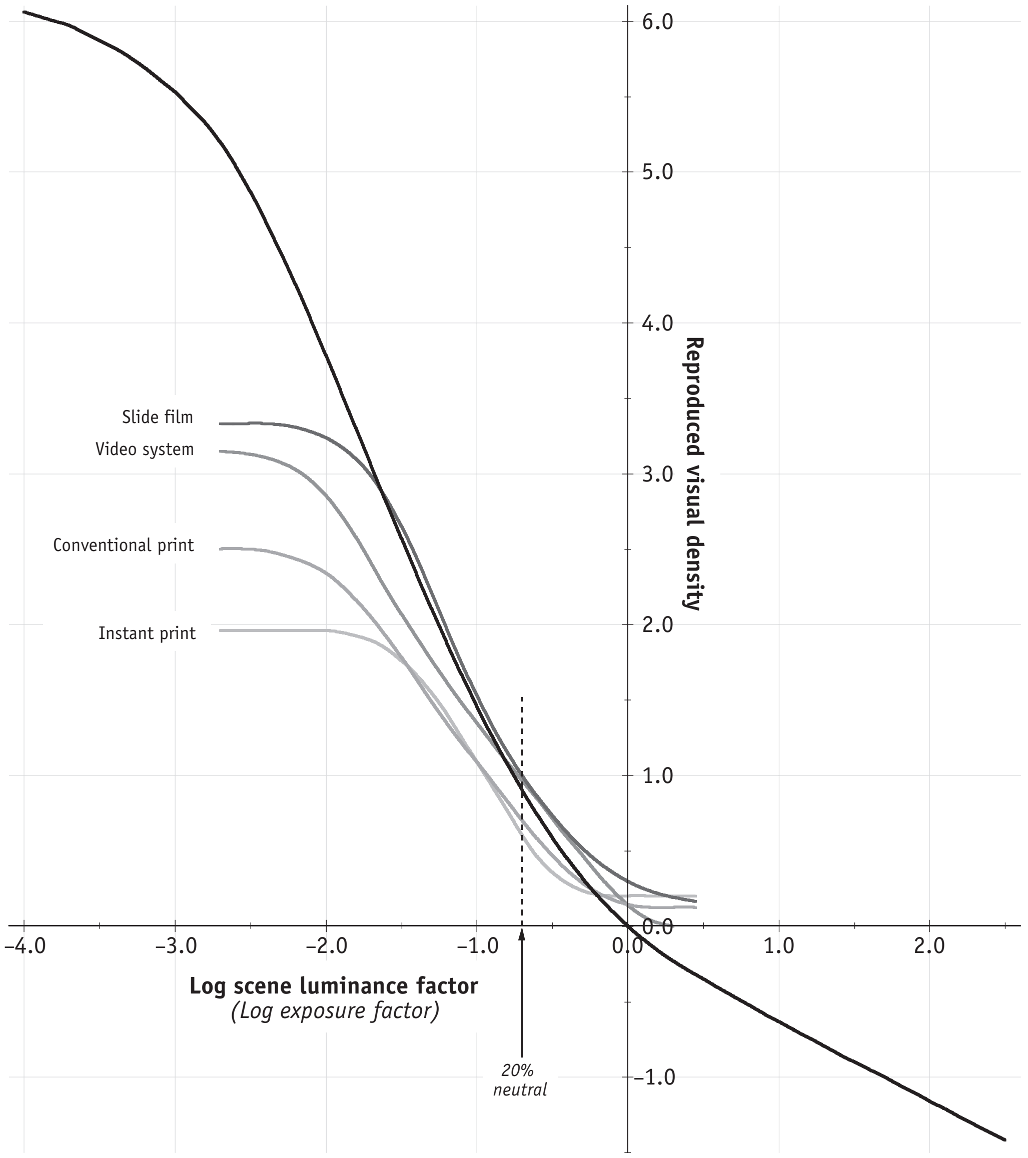


Figure 21.16 Grayscales for the RRT and four example systems, properly aligned along the relative log exposure axis. The alignment is based on the aim visual density that each should produce from a normally exposed reference. The grayscales also are properly aligned along the visual density axis such that they are matched for perceived overall brightness.

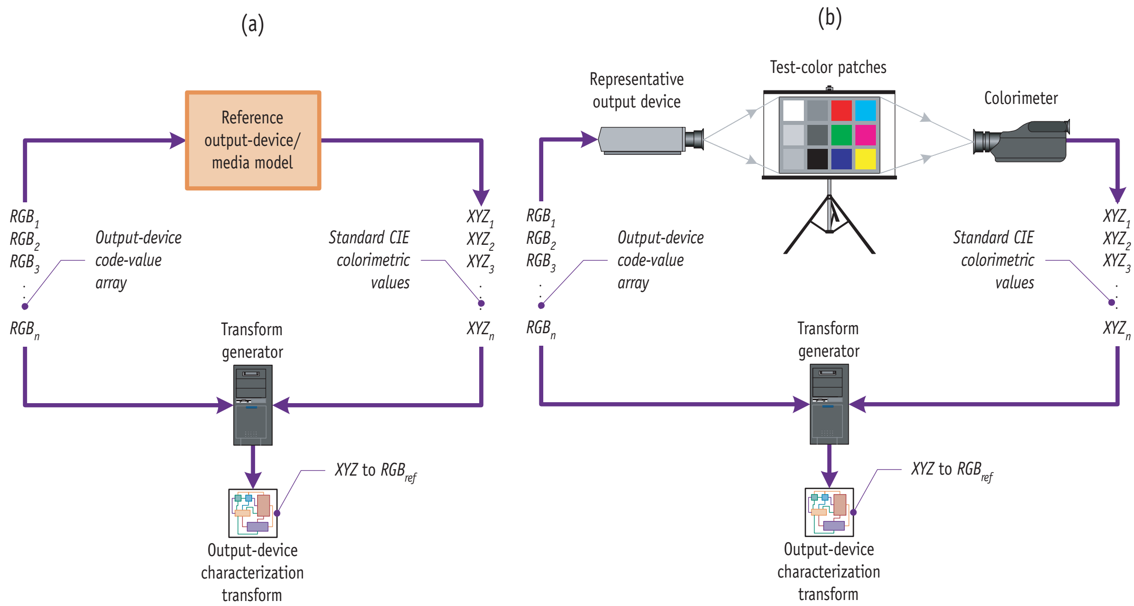


Figure 21.17 A colorimetric output-characterization transform can be developed from output device/media models (a) or from empirical data obtained by measuring colorimetric values of an appropriate number of color patches produced on the representative output device from an array of output-device code values (b).

perform. That is important because output calibration often must be performed on a frequent basis.

Digital cinema system implementation

As mentioned earlier, the example digital cinema system that has been described here is based on a proposal submitted to the Academy of Motion Picture Arts and Sciences. Accordingly, the example is not simply instructional—the system it describes is intended to be implemented. At the time of this writing, the Academy's Science and Technology Council is developing an ICES (currently called *ACES*, for *Academy Color Encoding Specification*) based on the input-image encoding principles described in the example. The same committee is also developing an RRT, again based on the example presented here.

An important consideration for any color-managed system intended for an industrial application is that it must provide continuity with existing systems and practices. As discussed earlier, the current interchange standards for digital cinema images are the Cineon DPX and ANSI/SMPTE DPX file formats. Because they are based on film printing-density values, DPX files can be readily input to the example system and transformed to ICES values by the same types of universal or product-specific transforms used with negative-film scanners. In addition, inverses of these transforms can be used to transform ICES values (from *any* input source, not just film) to universal or product-specific DPX files. These capabilities, which link the example system to existing systems and work flows, are indicated in Figure 21.2 by the two-way arrow connecting DPX files with the ICES.

The example system also is consistent with other aspects of digital cinema already in place. Specifically, the reference viewing conditions for the example OCES correspond to those currently specified for digital cinema projection, as defined in the SMPTE (Society of Motion Picture and Television Engineers) documents listed in the Recommended Reading at the end of this book. Other aspects of the example system are also consistent with current SMPTE objectives for digital cinema output. In particular, the SMPTE documents describe the need for a digital source master (DSM) image file from which various distribution elements (such as film distribution masters, digital cinema distribution masters, home video masters, and broadcast masters) can be made. In the example system presented here, a file of OCES values (or perhaps an ICES file plus the RRT) can serve the function of the DSM, and multiple ODTs relating OCES values to device values can be used to generate image data required for various types of required SMPTE distribution masters. Image archiving could be done by saving the DSM, the OCES file, or preferably the ICES and RRT.

Summary of key issues

- The flexibility and extensibility of the Unified Paradigm and its associated color-encoding method have been demonstrated in two complex example systems.
- Appearance-based color encoding was implemented in both systems by the use of appropriate colorimetric specifications and defined sets of encoding reference viewing conditions.
- These examples illustrate that with the use of appropriate color encoding specifications, powerful systems can be assembled by linking various system components and subsystems in parallel and serial arrangements.
- The objectives of the first example—a complex multiple input/output system—were achieved by the use of a single appearance-based color encoding specification (CES), parallel arrangements of the system's inputs and outputs, and the basic encoding principles described in Chapter 20.
- The objectives of the second example—a digital cinema system—were achieved by the use of a serial architecture composed of an input color encoding specification (ICES), a separate output color encoding specification (OCES), and a reference rendering transform (RRT) linking the two.
- The use of separate input and output encoding specifications in the digital cinema application allows each CES to be optimized based on its specific requirements.

22
A Unified Paradigm: Color Interchange

In Chapters 20 and 21, the principles of the Unified Paradigm were applied to various types of individual color-imaging systems. In this chapter, those same principles will be applied to our ultimate goal—a comprehensive, unified color-management environment for the entire color-imaging industry. The unifying factor will be a method for the unambiguous and unrestricted communication of color among all systems operating within the environment.

Figure 22.1 illustrates a unified color-management environment in which various types of systems are operating. Some systems (the *ABC* systems) are independently providing all functions of the Unified Paradigm. Other systems and groups of related systems are operating according to just one or two of the Type *A*, *B*, or *C* paradigms. That is as it should be. There are valid reasons— simplicity, for one—to limit the functionality of particular systems, and a truly unified environment must include and fully support such systems.

Similarly, different color-encoding methods and color-encoding data metrics might be used internally on particular systems or for color interchange among groups of related systems operating within the overall environment. As was discussed earlier, there are important engineering reasons for using different color-encoding methods and data metrics on different types of imaging systems. A color-management environment that excluded systems based solely on their use of particular encoding methods and/or data metrics would not truly be unified.

Figure 22.1 also shows multiple forms of color interchange, with different system-specific color encoding specifications, being used between pairs of related systems within the overall environment. For example, two identical imaging systems might be directly interchanging color information encoded in terms of system-specific densitometric or colorimetric values. They might also be using a particular type of device-dependent color-interchange method. Those forms of interchange are unambiguous when used between pairs of systems that are based on the same color encoding specification. A unified color-management environment must allow these direct, system-specific forms of color interchange because they generally are fast and efficient. In some cases, they may provide the only practical form of interchange for large amounts of color data between systems of the same type.

However, the ultimate success of a unified color-management environment also requires a *general* form of color interchange (indicated by dashed arrows in the figure) that can be used for communicating color among dissimilar, as well as

Digital Color Management: Encoding Solutions, Second Edition Edward J. Giorgianni and Thomas E. Madden
 ISBN: 978-0-470-51244-9

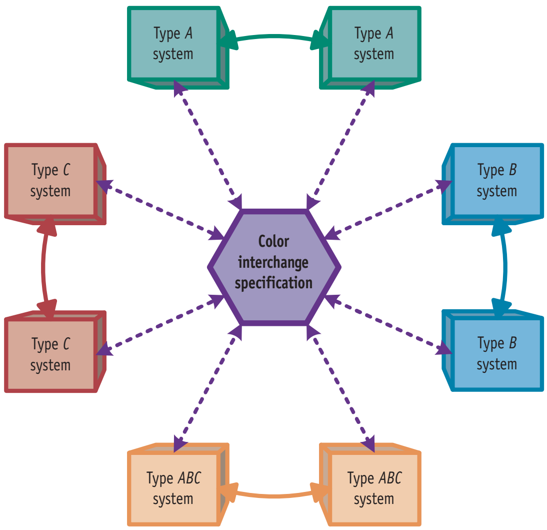


Figure 22.1 A unified color-management environment. Some systems operating within this environment are providing all the functions of the Unified Paradigm. Others are operating according to various system-specific restrictions. All systems can communicate with each other through a color interchange specification (CIS).

related, types of imaging systems operating within the environment. As with all successful color communication, this communication must be unambiguous. The receiving system should not need to seek additional information in order to interpret an interchanged color. General color interchange also must be unrestricted. The method on which it is based must not impose limitations on luminance dynamic range, color gamut, or color interpretation.

These criteria are identical to those determined earlier as being essential for the *internal* input-to-output communication of color for an individual color-imaging system that is independently capable of providing all the functions of the Unified Paradigm. This is a logical outcome since a single system having such all-inclusive capabilities encompasses all other systems having various subsets of those capabilities. Because the criteria are the same, the same appearance-based solution that was used for the internal color encoding of a Unified Paradigm system can be used for general color communication among all systems.

This solution can be expressed in terms of what will be referred to as a *color interchange specification* (CIS). A fully defined CIS must include:

- A complete colorimetric specification of the color to be communicated.
- A fully defined data metric in which to express that colorimetry.
- A specified set of interchange reference viewing conditions.

A color-imaging system can participate in this general color interchange if (and *only* if) its own internal color encoding is defined *completely*. The color encoding specification for a system's internal encoding, like a CIS, must include complete descriptions of its colorimetry, data metric, and associated viewing conditions. This degree of specificity is required in order for meaningful conversions to be made from a given system's internal encoding to the interchange specification for outgoing color communication, and from the interchange specification to a system's internal encoding for incoming color communication (Figure 22.2).

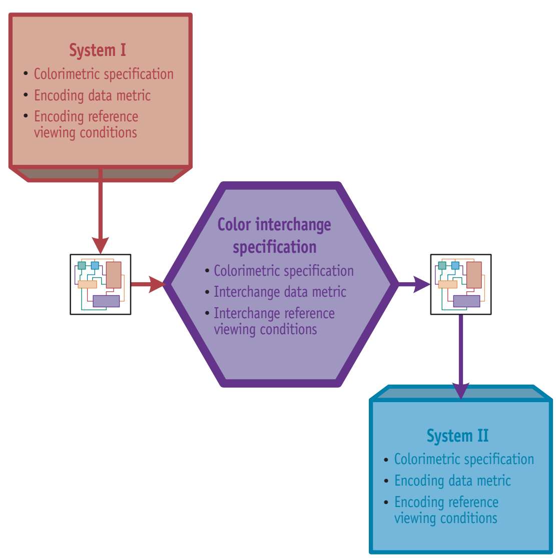


Figure 22.2 Use of a color interchange specification for color communication between two imaging systems.

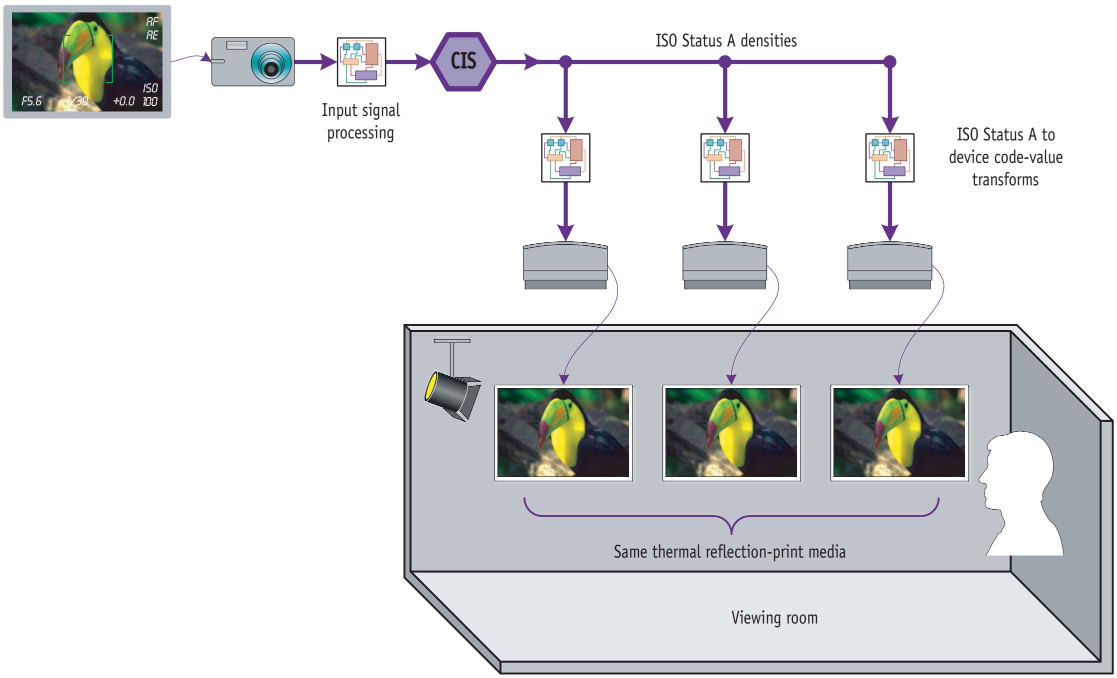


Figure 22.3 An example of system-specific color communication. This type of communication can work successfully, but only if certain rules are adhered to.

Providing a complete definition of the internal color encoding of a system requires an examination of how that system is used in practice. That can be an interesting exercise. In many cases, it will be found that a system has operated well only because certain rules were *implicitly* adhered to, even though those rules may never have been specified formally. For unambiguous color communication to and from other systems, such rules must be defined *explicitly*.

For example, one system that has been used for many years at a particular site uses digital cameras and several thermal printers to produce reflection prints. In that system, colors are encoded and communicated to the thermal printers in terms of Status A reflection density values (Figure 22.3).

That method of color interchange has worked quite well, but only because two unwritten rules always are followed: (1) the same thermal medium always is used on each of the printers; and (2) the resulting output prints always are examined in the same viewing room.

For communication with other types of imaging systems in a unified color-management environment, these rules would have to be defined explicitly so that encoded Status A density values can be translated unambiguously to a specification of color appearance. That color-appearance specification then can be expressed in terms of color-interchange values defined according to the CIS.

The definition of the Status A CIS and the definition of the CIS are compared in Table 22.1. Once these definitions are completed, a translation between the two color representations can be developed. That process is discussed next.

The first step in this translation process is to convert the encoded Status A density values to standard CIE colorimetric values. The conversion can be performed using a model of the illuminated thermal medium. The model would include the spectral characteristics of the medium's image-forming dyes, the optical characteristics of the reflection support, and the measured spectral power distribution of the viewing-room illuminant. The conversion also could

Table 22.1 Comparison of a densitometric CIS to an example CIS of a unified environment.

	Example System	Example CIS
Colorimetry	ISO Status A	1931 CIE
Data metric	Status A density × 100	24-bit CIELAB
Viewing conditions:		
Viewing flare	1.0 %	1.0 %
Image surround	Average	Average
Luminance level	150 cd/m^2	60–160 cd/m^2
Adaptive white	$x = 0.3303, y = 0.3395$	$x = 0.3457, y = 0.3585$

be derived empirically, using the method illustrated in Figure 22.4.

The chromaticity coordinates of the viewing-room illuminant can be computed from its spectral power distribution ($x = 0.3303$, $y = 0.3395$ in this example). These values can be used to determine an

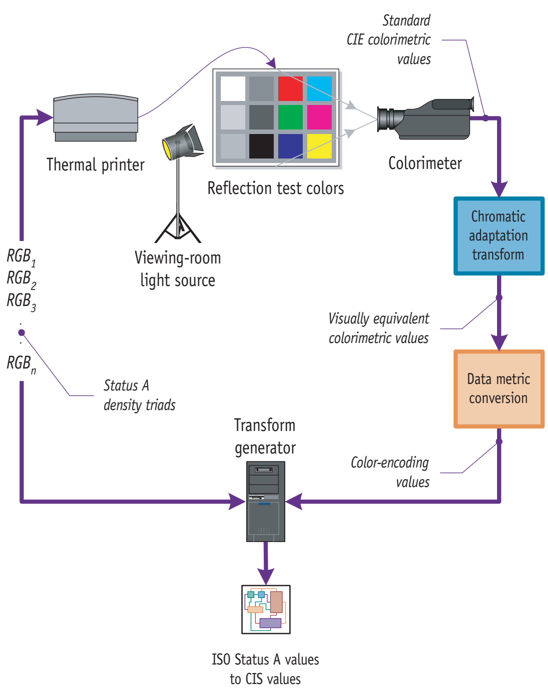


Figure 22.4 A transform to convert Status A density values to CIE colorimetric values can be derived from colorimetric measurements of test colors generated from encoded Status A values.

appropriate chromatic adaptation transformation, from the viewing-illuminant adaptive white to the D_{50} adaptive white of the CIS.

No other colorimetric transformations are required because the other conditions conform to the interchange specifications: the room illumination is such that the luminance of a white in a viewed image typically is about 150 cd/m^2; the viewing-flare luminance is approximately 1 % that of a white in a print; and the prints always are viewed on a neutral gray surface of about 20 % reflectance, which is consistent with an average-surround condition.

Many commercial systems have operated successfully under similar sets of tacit rules. For example, graphic arts images frequently are interchanged in terms of CMYK values. That interchange has worked reasonably well because what it really has meant is "CMYK values, for a particular assumed set of inks, printed onto an assumed substrate, for images to be viewed under a particular assumed set of conditions." As long as everyone involved in the interchange process understands and follows the same rules, this specialized form of color interchange can be sufficient. For general color interchange, however, CMYK values would have to be translated to colorimetric values, and the image-viewing conditions would have to be defined explicitly so that appropriate color-appearance transforms can be derived.

Similarly, digital images often are interchanged in terms of *RGB* video code values. That also can work without additional transformations if everyone involved uses similar display devices and viewing conditions. But again, for general color interchange, the *RGB* values must be translated to colorimetric values. This can be done by specifying the characteristics of a reference display device. In addition, a set of reference display viewing conditions must be defined explicitly.

Interchange transforms

An interchange transform can be developed for any color-encoding method and data metric that can be *completely* specified in terms of the factors listed in Table 22.1. The transform must account for any differences in the methods used in making colorimetric measurements, for any differences in the data metrics, and for any differences in the viewing conditions associated with the color-encoding method and the CIS involved.

In cases where the color encoding specification (CES) of a color-imaging system is identical to the CIS of the color-management environment, no transformation is needed between the two. In most cases, however, transformations will be required. For example, Table 22.2 compares the specifications for a particular scene-space CES to an example CIS. The basic steps involved in the required transformation are shown in Figure 22.5.

- Data metric conversion from digital code values CV_1, CV_2, and CV_3, as defined by the data metric of the scene-space CES, to (scene space) CIE XYZ tristimulus values.
- Adjustment of the XYZ values to compensate for the greater level of viewing flare in the CIS reference viewing conditions.
- Further adjustment of the XYZ values to compensate for the overall lower luminances of images in the CIS reference viewing conditions.
- Further adjustment of the XYZ values to account for the difference in the observer adaptive white chromaticity between the CES and CIS, from a chromaticity of D_{65} to a chromaticity of D_{50}.
- Metric conversion of the adjusted XYZ values to CIELAB L^*, a^*, and b^* values.
- Scaling of the CIELAB L^*, a^*, and b^* values to form digital code values, CV_1, CV_2, and CV_3, as defined by the data metric of the CIS.

Discussion

Interchange transforms based on a completely defined, appearance-based color interchange specification do not alter the color appearance represented by the color values being transformed, nor do they limit or restrict the communication of color in any way. These properties are essential for the successful implementation of the Unified Paradigm. They allow individual systems to use various color-encoding methods and data metrics that are most appropriate for their particular requirements,

Table 22.2 Comparison of an example scene-based CES to an example CIS of a unified environment.

	Example Scene-Based CES	Example CIS
Colorimetry	1931 CIE flareless measurement	1931 CIE flareless measurement
Data metric	CIEXYZ	CIELAB
	$X = 200\,(CV_1/1023)$	$CV_1 = 2.10L^*$
	$Y = 200\,(CV_2/1023)$	$CV_2 = a^* + 128$
	$Z = 200\,(CV_3/1023)$	$CV_3 = b^* + 128$
Viewing conditions:		
Viewing flare	None	1.0 %
Image surround	Average	Average
Luminance level	>1600 cd/m^2	60–160 cd/m^2
Adaptive white	$x = 0.3127$, $y = 0.3290$	$x = 0.3457$, $y = 0.3585$

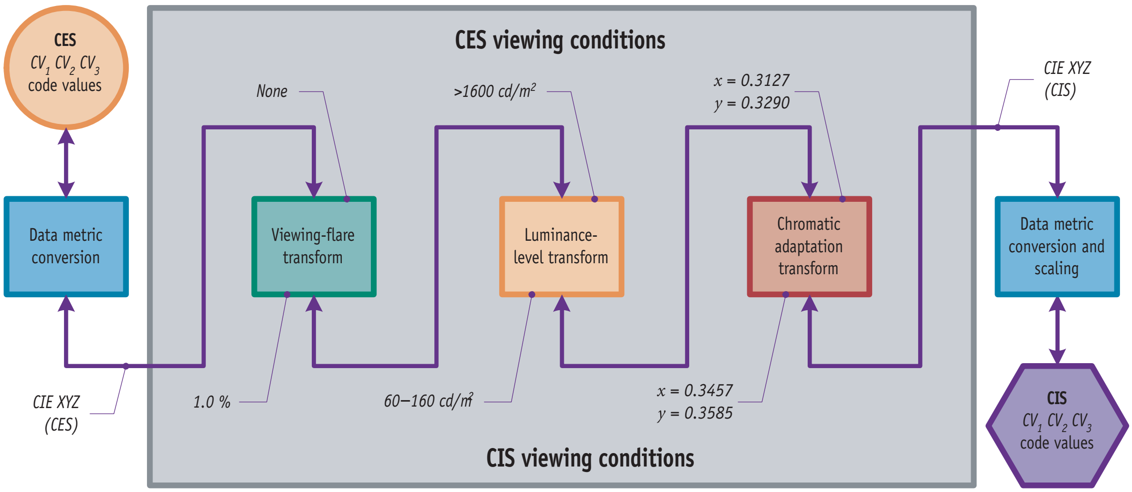


Figure 22.5 Transformation between an example scene-space-based CES and an example CIS.

yet they also provide a means by which all systems can communicate unambiguously with one another.

With the addition of this solution to the problem of general color interchange, nearly everything required for a completely unified color-management environment is in place. The last requirement—a practical method of implementation—is the topic of the next chapter.

Summary of key issues

- A unified color-management environment must be capable of supporting restricted (as well as unrestricted) color-imaging systems operating within that environment.
- A unified color-management environment must support the internal use of various color-encoding methods and data metrics on systems operating within that environment.
- A unified color-management environment must support the use of various system-specific forms of color interchange among like systems.
- A unified color-management environment also requires a general method for interchanging color among all types of imaging systems operating within the environment.
- General color interchange must be unambiguous and unrestricted; the interchange method itself must not impose limitations on luminance dynamic range, color gamut, or color interpretation.
- In the unified color-management environment, general color interchange among systems is based on a color interchange specification (CIS). The CIS includes a complete colorimetric specification, a fully defined data metric, and a specified set of interchange reference viewing conditions.
- For participation in general color interchange, the internal color encoding of a given system also must be completely defined in terms of its colorimetric specification, data metric, and associated viewing conditions.
- When appropriate techniques are used, interchange transformations retain the color appearance represented by the color values to be transformed.

23

A Unified Paradigm: Implementation

In this chapter, a number of alternative implementations of the Unified Paradigm will be discussed in a context based on one of several possible overall system architectures. The example architecture is straightforward and common to many existing color-management systems. However, while it can *support* the implementation of the Unified Paradigm, no architecture alone can *create* a Unified Paradigm system. As discussed in previous chapters, the properties and functionality of the Unified Paradigm result from its underlying philosophy and its methodology of representing color.

General architecture and components

The architecture that will be used in this chapter incorporates two basic components. The first component, a *profile*, contains a digital signal processing transform (or a collection of transforms) plus additional information concerning the transform(s), the device, and the data. Input profiles and output profiles provide the information necessary to convert device color values to and from values expressed in terms of a defined color encoding specification (CES). In any practical implementation, detailed format standards for these profiles must be available so that profiles from different providers can be used interchangeably. The second basic component, a *color-management module* (CMM), is a digital signal processing "engine" for performing the actual processing of image data through profiles. A CMM also may perform other computational tasks, as will be described shortly.

Figure 23.1 shows a simple arrangement for connecting a single input device to a single output device, using an input profile, an output profile, and a CMM. Figure 23.2 shows a more complex multiple input/output arrangement.

In an overall color-managed environment based on the use of profiles and CMMs, color-image data can be interchanged unambiguously among imaging systems in three different ways, as illustrated in Figure 23.3. These three alternatives are as follows:

- Image data can be processed through appropriate input profiles and then interchanged directly in terms of CES values.
- Image data can be interchanged in terms of input-device values, and an appropriate input profile relating input-device values to the CES can be embedded in the image file for later use by a CMM.
- Image data can be interchanged in terms of input-device values, but without an embedded input profile. An image file header or other form of metadata then would include an identifier that can be used

Digital Color Management: Encoding Solutions, Second Edition Edward J. Giorgianni and Thomas E. Madden
 ISBN: 978-0-470-51244-9

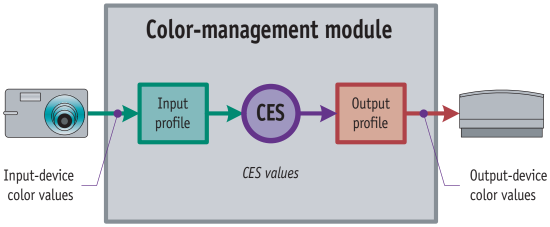


Figure 23.1 Connection of a single input device to a single output device, using an input profile, an output profile, and a color-management module (CMM).

by a CMM to select a particular input profile from a stored library of profiles.

Although all three alternatives are feasible, the first method may involve an additional data quantization step that can produce undesirable image artifacts. That is one reason why the second alternative is more commonly used. Another reason why the second alternative often is preferred is that since an image being received has not yet been processed through an input profile, the recipient can choose to apply a different input profile if desired. The third alternative essentially provides the same features as the second. However, it also requires some agreed-upon method for administering the registration of profiles and profile identifiers.

Overall architecture

One basic concept for an overall architecture is shown (Figure 23.4). At the top of the architecture are user applications. These applications are linked to graphics libraries and imaging libraries through an application programming interface (API). The API in turn is linked to a color-management framework. Connected to this framework are individual profiles, which conform to profile format standards, and CMMs. In most cases, a default CMM would be provided by a computer operating system. The architecture also allows for other (perhaps more advanced) CMMs to be provided by other parties. This architectural arrangement provides considerable flexibility, and it allows for future expansion of the overall structure.

There are two ways that a CMM operating in this architecture might process an image. It might do so by first processing input-device color values through an input profile to form CES values and then processing those values through an output profile to form output-device color values (Figure 23.5a). A CMM instead might first generate a concatenated composite profile from the separate input and output profiles and then process input-device color values through the composite profile (Figure 23.5b). The two techniques will yield similar results, although the latter generally is more efficient. It also avoids

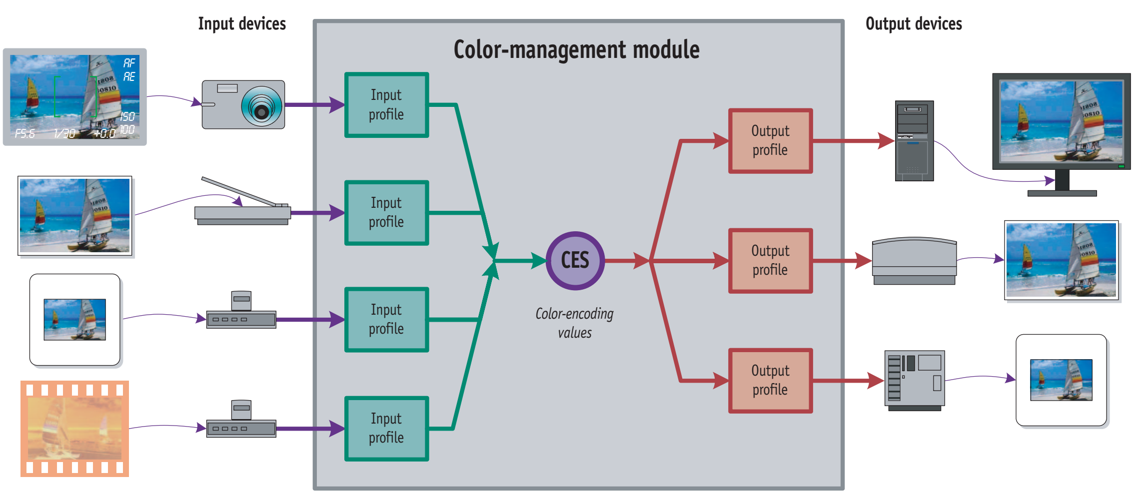


Figure 23.2 A more complex imaging system comprising multiple input profiles, multiple output profiles, and a CMM.

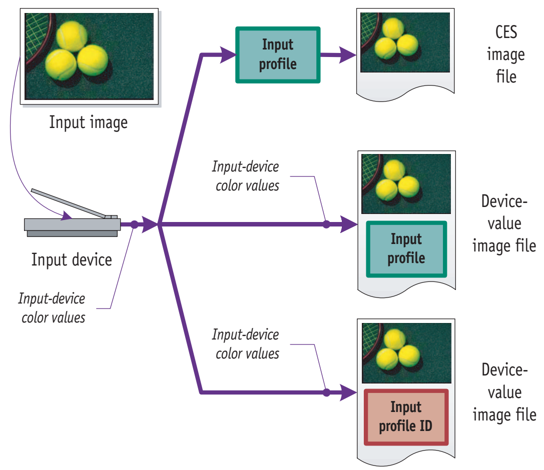


Figure 23.3 Digital images can be interchanged among systems in three different ways.

quantization errors that can occur when CES values are digitized.

The overall architecture also allows the use of other types of profiles. These profiles can be used to perform various types of transform operations such as data metric conversions. They also can be used to implement other transformations specified from a user application. For example, an imaging application might create a profile as a means for producing an overall color-balance adjustment to an image. The image then would be processed through the input profile, transform profile, and output profiles to generate color-balanced output-device values (Figure 23.6a). Alternatively, a composite profile generated by a concatenation of the three separate profiles could be used (Figure 23.6b) to perform the entire signal processing.

Implementations of the Unified Paradigm

There are several approaches that can be considered for implementing the unified color-management paradigm in an environment based on the described architecture. The first, illustrated in Figure 23.7, conforms most closely with the general approach used in previous chapters. The input and output

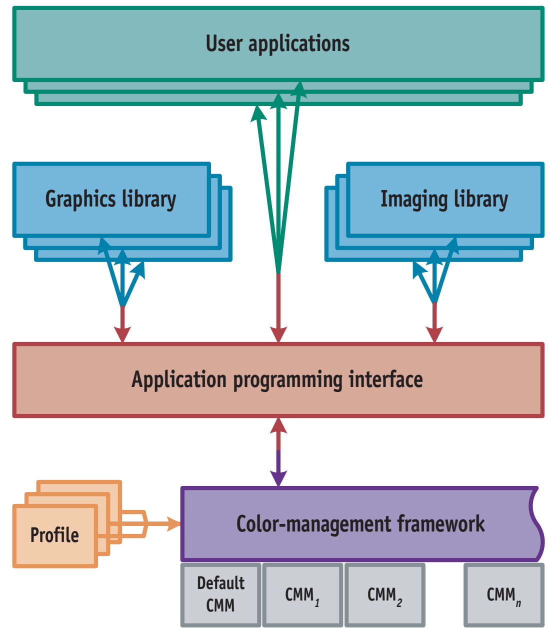


Figure 23.4 A basic system architecture for a color-managed environment.

profiles perform all of the input/output transformations (viewing-flare compensations, adaptation adjustments, gamut mappings, etc.) that have been discussed. Other transformations, specified according to various user-selected input options, also are implemented in the profiles. That is why *multiple* input profiles are shown for each input—each

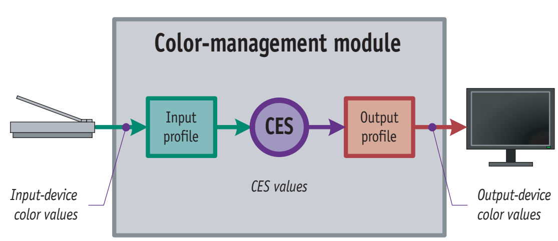


Figure 23.5a One method by which a CMM might process an image. Image values first are processed through an input profile to form CES values that then are processed through an output profile to form output values.

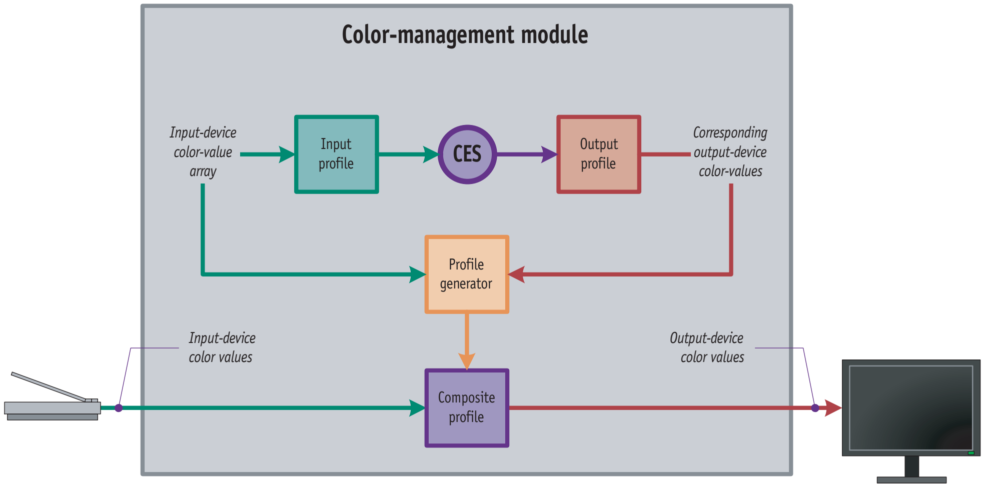


Figure 23.5b Another method by which a CMM might process an image. The CMM first generates a composite profile from the separate input and output profiles and then process image values through the composite profile.

profile in the stack represents a different user-specified interpretation of input color. The different output profiles for a given output device similarly represent various user-selected output options, such as alternative re-rendering or gamut-mapping strategies.

In this implementation, the CES functions as both a CES within any given system and as a CIS for the overall environment. In Figure 23.7, for example, the CES serves as a CES to connect each system input to its own output, and it serves as a CIS to connect any input of one system to any output of another. The principal feature of this implementation is that a CMM can be very simple; its only function is to provide a direct connection between input and output profiles. No additional image data identification is required because the input profiles themselves serve to create compatibility among all the inputs.

A second approach, shown in Figure 23.8, uses simpler input/output profiles. In this implementation, the profiles might represent transformations to and from standard CIE colorimetric values. This approach requires the use of metadata or some other means of providing information concerning the type

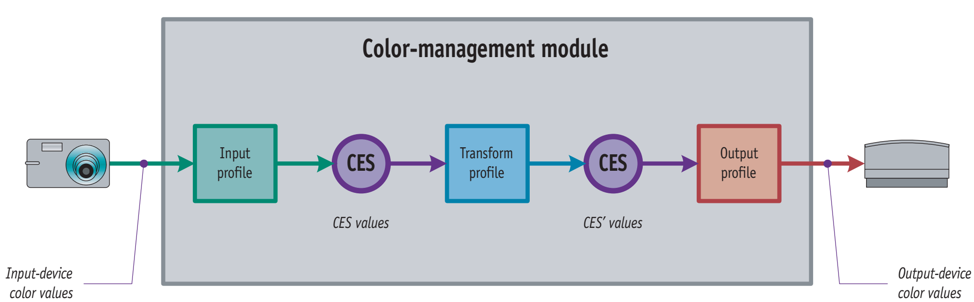


Figure 23.6a Use of a transformation profile in a signal processing chain.

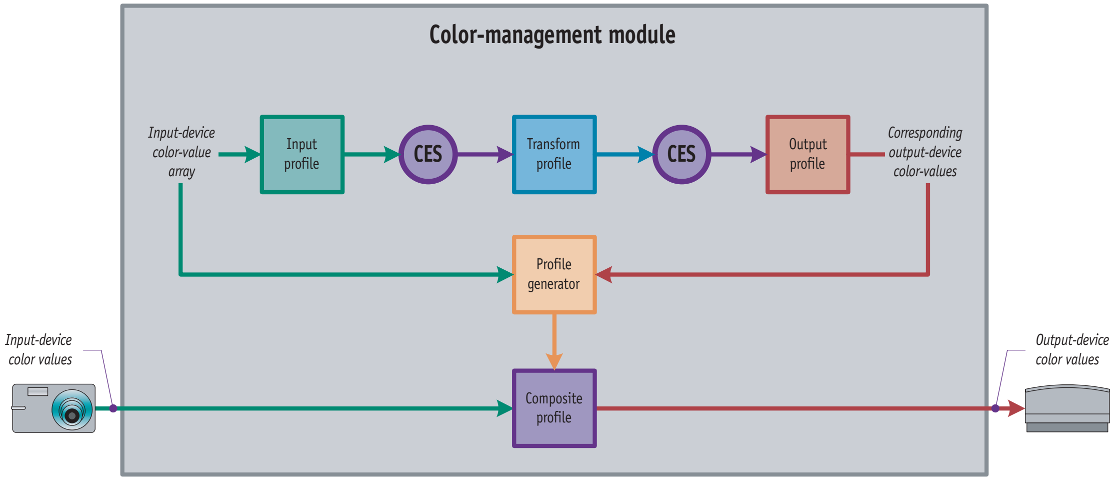


Figure 23.6b Derivation and use of a composite profile that includes an additional transformation profile.

of device, medium, viewing environment, and so on. In addition, it requires a somewhat more advanced CMM that can use that information, together with the information provided by the user's responses to the input and output options of the Unified Paradigm, to determine the required signal processing transformations. For example, suppose the input medium is identified as being a photographic slide film, the output is to be an inkjet reflection printer, and the user option is to match the output to the input. The CMM computations would have to include appropriate colorimetric transformations to account for the different viewing conditions associated with the input and output media.

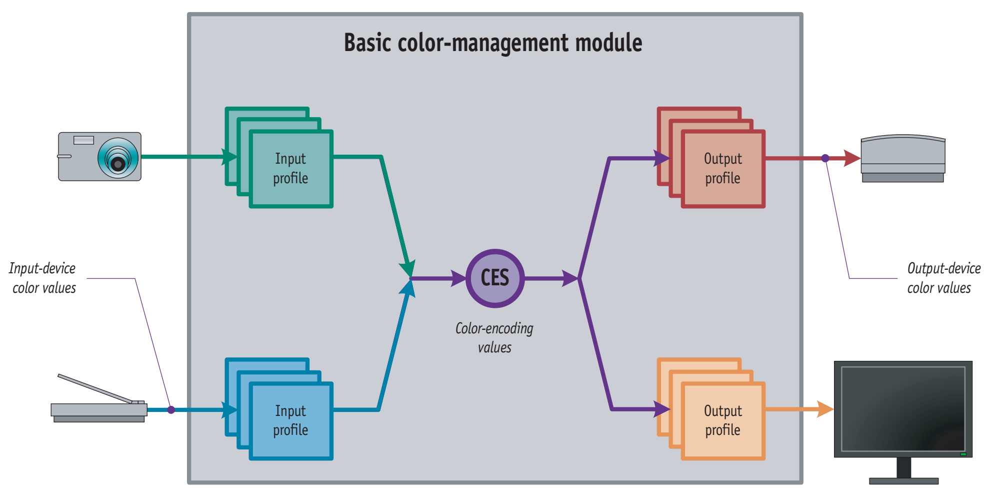


Figure 23.7 In this implementation, a single CES functions as a CES within a given system and as a CIS for the overall environment. Since any difficult color transformations are contained in the profiles, the CMM can be basic.

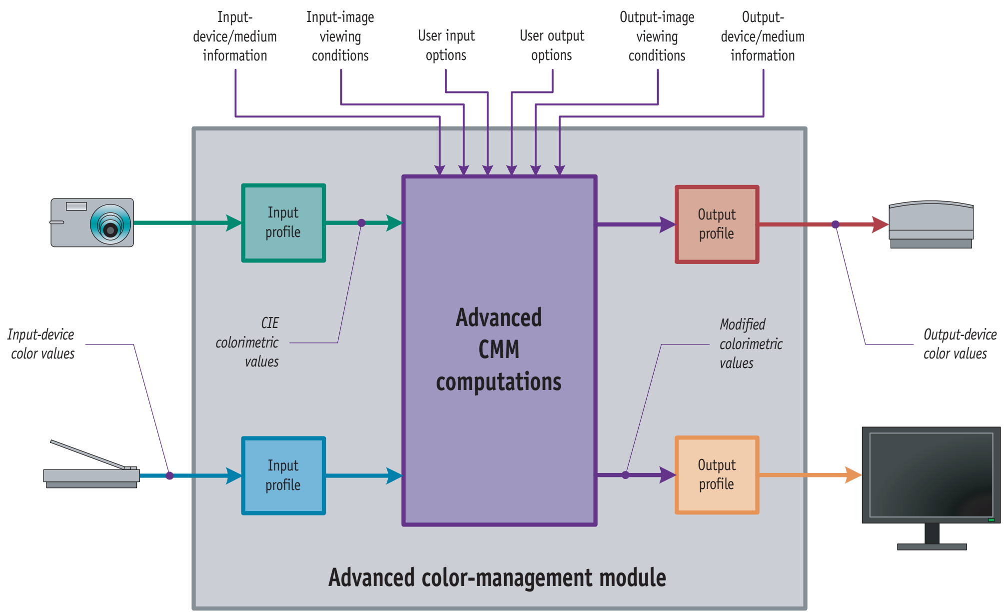


Figure 23.8 In this implementation, the profiles are simpler, and the more difficult transformations are handled by a more advanced CMM.

The CMM computations also could provide appropriate output color-gamut mapping, although that might instead be included in the output profiles.

Figure 23.9 shows a third approach. In this implementation, simpler input and output profiles again are used. But instead of the appearance transformations and other difficult transformations being handled entirely by a CMM, they are handled either by the application alone or by some combination of the application and a CMM. This approach, like the second approach, requires metadata or some other means for providing information to the application concerning factors such as input/output device types, media, viewing conditions, and color-interpretation options. In the particular implementation shown in Figure 23.9, that information is used to first construct a transformation profile, which then is used by a basic CMM.

Each of these approaches has important advantages, and each warrants consideration. However, we believe other approaches that *combine* the best features of the three described approaches also should be considered. We suggest that particular consideration be given to combined approaches in which the color encoding of the CES is defined such that excellent results can be obtained using basic CMMs and user applications. These basic CMMs and applications would only need to provide direct connection of input, transformation, and output profiles. In addition, however, our proposed approaches would incorporate a file format that includes metadata. Information contained in that metadata would be used by more advanced CMMs and applications that optionally can go beyond the direct connection of profiles. For example, an advanced CMM or application might derive an output profile incorporating a highly customized gamut-mapping transformation, based on knowledge of the specific input medium, output medium, and user preferences. Again, however, the CES would be defined such that advanced CMMs and/or applications are not mandatory.

One possible combined approach is shown in Figure 23.10a (and also in Figures 23.10b and 23.10c). The basic approach is most like that of the first implementation, which was shown earlier in Figure 23.7, in that the CES is acting as an appearance-based CES

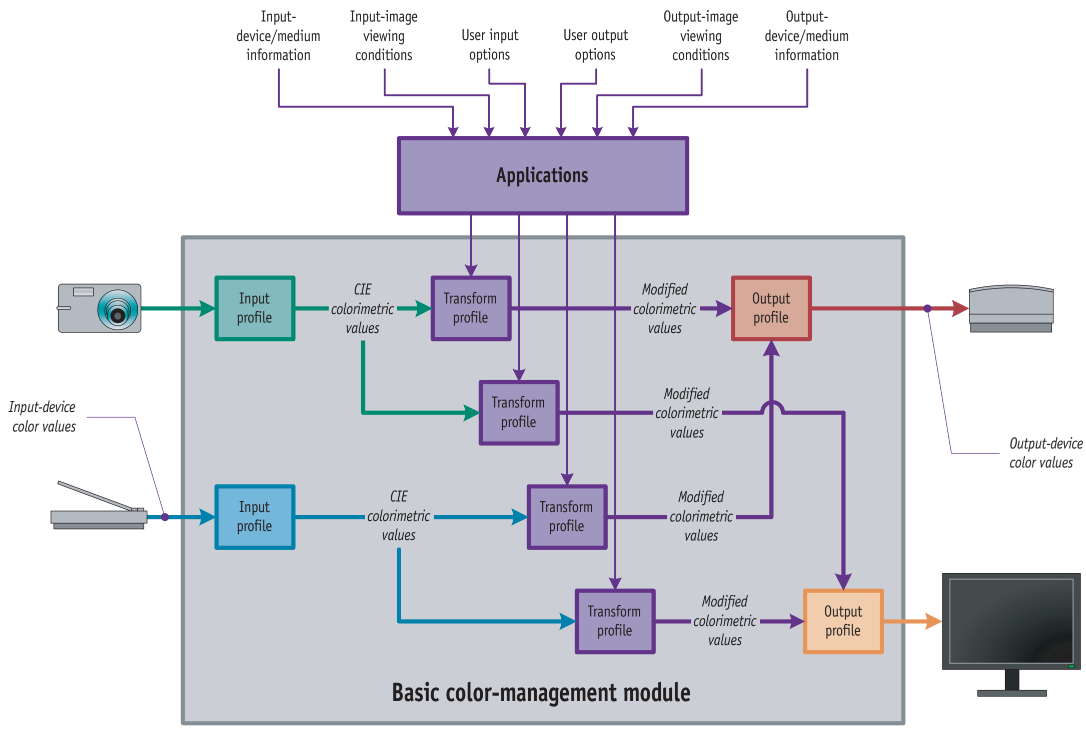


Figure 23.9 In this implementation, a basic CMM can be used because the more difficult color transformations are handled by the application.

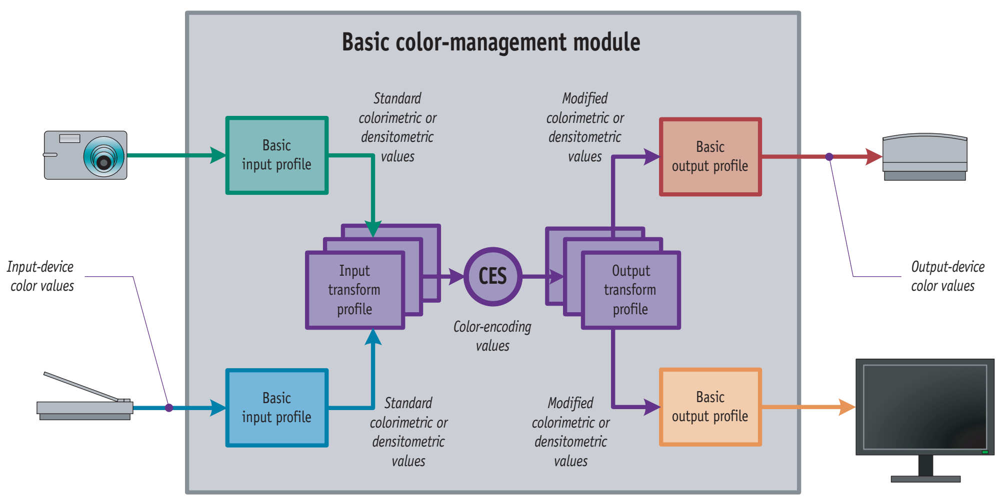


Figure 23.10a In this split-profile implementation, input and output profiles convert between standard colorimetric or densitometric values and device values. Additional profiles are used to convert between those standard values and CES values.

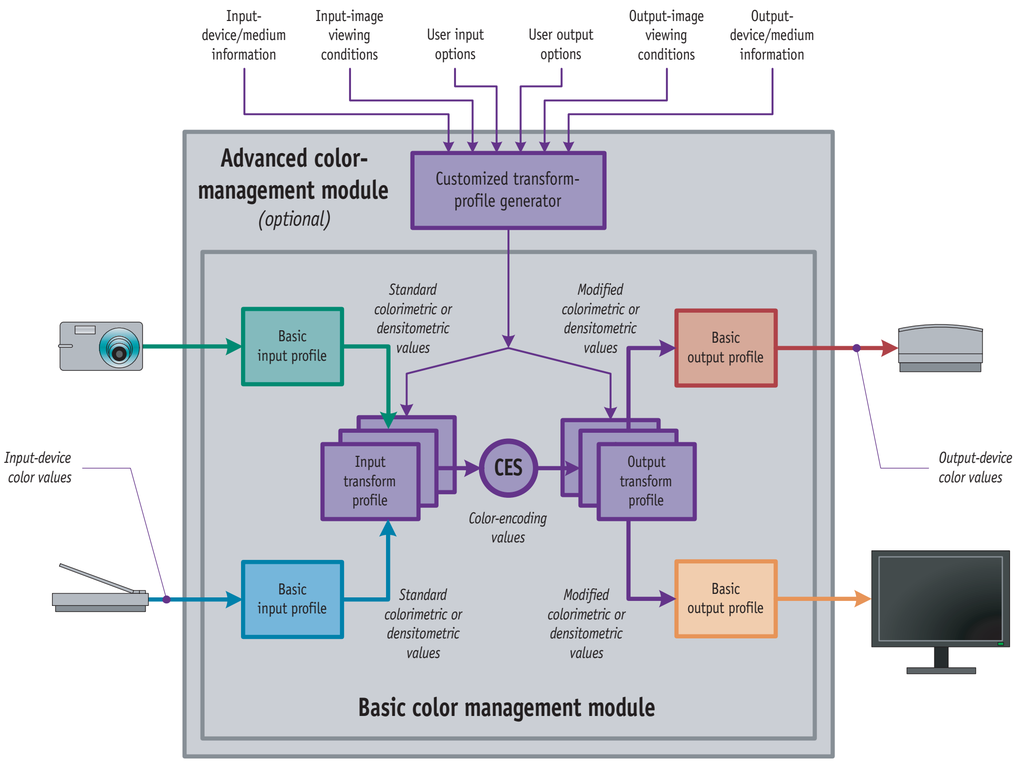


Figure 23.10b Use of metadata information by an advanced CMM in a split-profile implementation. In this implementation, an advanced CMM derives customized input and/or output profiles based on the information provided in the metadata.

and CIS. Note, however, that the input and output profiles have been *split* into two parts.

Input profiles convert input-device values to standard colorimetric or densitometric values, and output profiles convert from standard colorimetric or standard densitometric values to output-device values. Additional profiles, containing the more complex color-appearance transformations, then are used to convert between standard colorimetric or densitometric values and CES values. Figure 23.10b shows a similar arrangement and the optional use of metadata information by an advanced CMM. That CMM derives customized input and/or output profiles based on the information provided in the metadata.

One advantage of a split-profile approach is that it allows direct communication of system-specific color values among related devices, as shown in Figure 23.10c. For example, two related devices might communicate in terms of printing-density values, monitor RGB values, or other system-specific values. In some implementations, this direct communication can offer greater signal processing speed and accuracy.

Perhaps the most important advantage of the split-profile approach is that it greatly simplifies the task of equipment manufacturers and suppliers who wish to provide profiles for their devices. For example, a scanner or digital camera manufacturer only would need to provide a relatively simple input profile for transforming input-device values to standard colorimetric or densitometric values, rather than a more complex profile that also included color-appearance transformations. Similarly, the

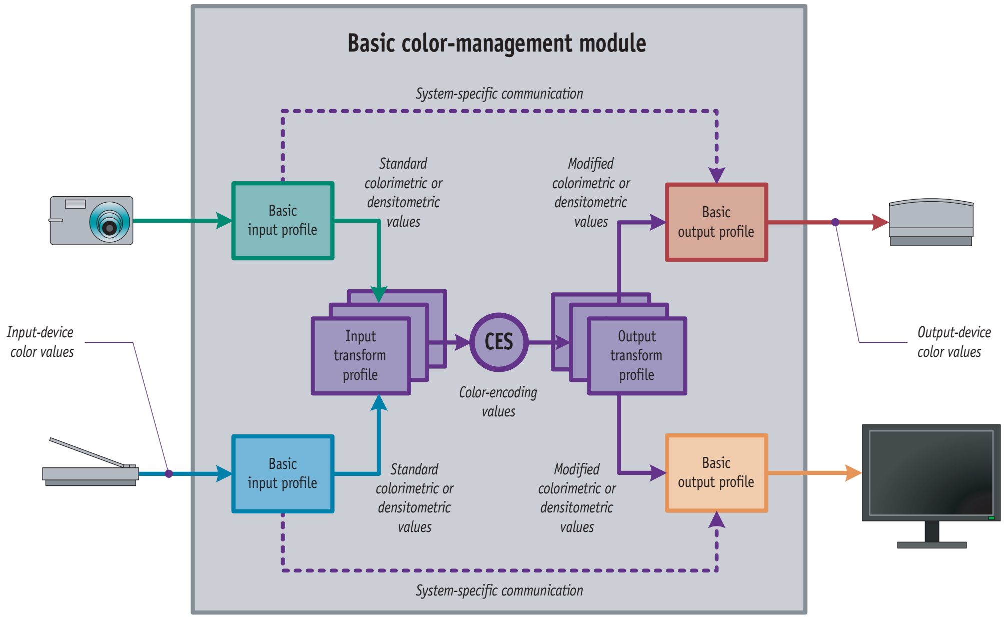


Figure 23.10c In a split-profile implementation, system-specific color values can be communicated directly (indicated by dashed lines) among related devices. In some implementations, such communication offers faster signal processing speed and greater accuracy.

manufacturer of an output device only would need to provide a relatively simple output profile for transforming standard colorimetric or densitometric values to output-device code values. Profiles containing the more complex transformations then could be supplied by others, such as color-management software developers.

In practice, it generally would be useful to further divide the input and output profiles into separate *characterization* and *calibration* profiles, as illustrated in Figure 23.11a. This arrangement is particularly advantageous in imaging systems having multiple input and/or multiple output devices of the same basic type. As explained in Chapter 14, the advantage is that a single (and generally more complex) characterization profile can be used for a number of related devices. Each individual device then only would need a relatively simple calibration profile. Figure 23.11b illustrates this use of multiple calibration profiles, with related characterization profiles, in a split-profile implementation.

In some situations, it would be more advantageous for the calibration profiles to reside in the input/output devices themselves (Figure 23.11c). For example, in a system that uses many output devices of the same basic type, output-image files could be produced by the CMM without requiring knowledge of the specific output device that will be used. Image data then could be sent to any of the output devices because, in a calibrated state, each device matches the characteristics of the reference device for which the characterization profile was derived.

Another advantage, shared by all of these split-profile implementations, is that updates can be implemented easily. For example, the technology for performing appearance transformations no doubt will continue to be improved. With split profiles, implementing improvements could be accomplished simply by updating a relatively small number of colorimetry-to-CES profiles and CES-to-colorimetry profiles. By comparison, *all* input and

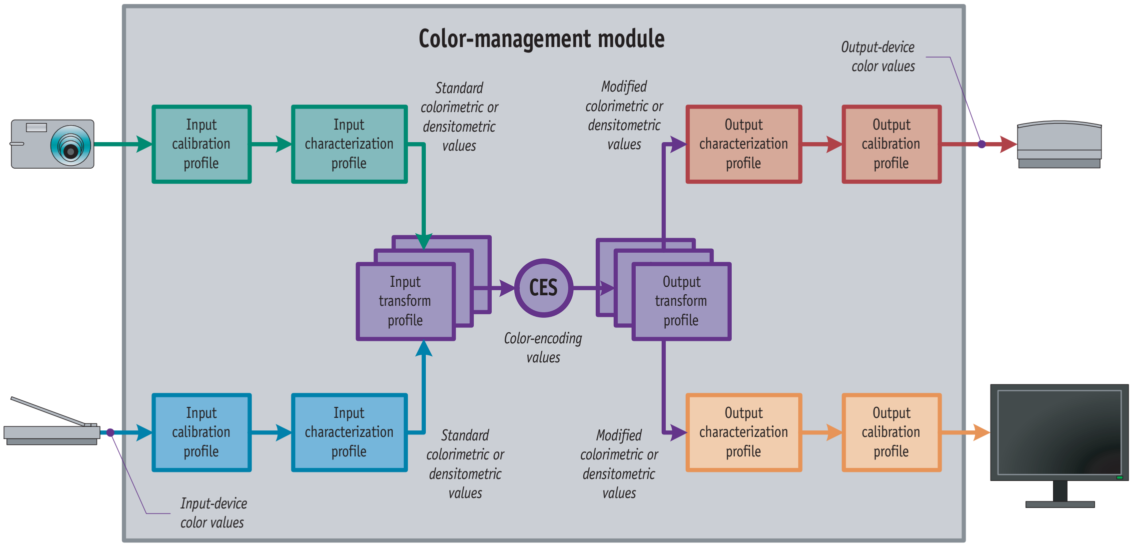


Figure 23.11a In this split-profile implementation, the input and output profiles are further divided into characterization and calibration components.

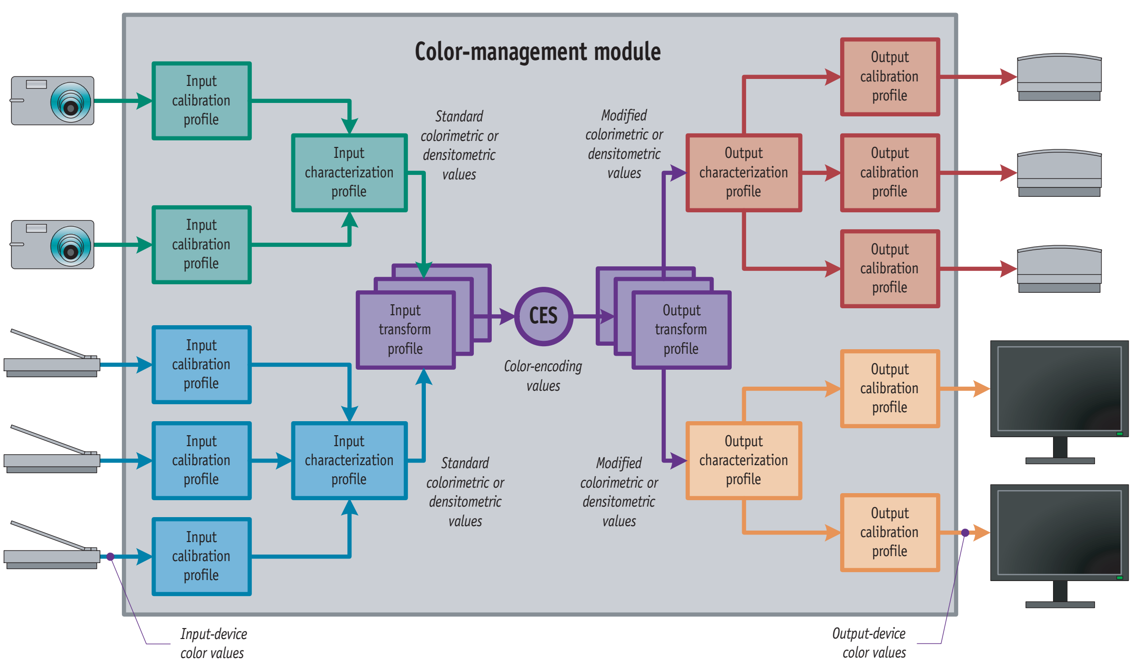


Figure 23.11b Use of multiple calibration profiles, with related characterization profiles.

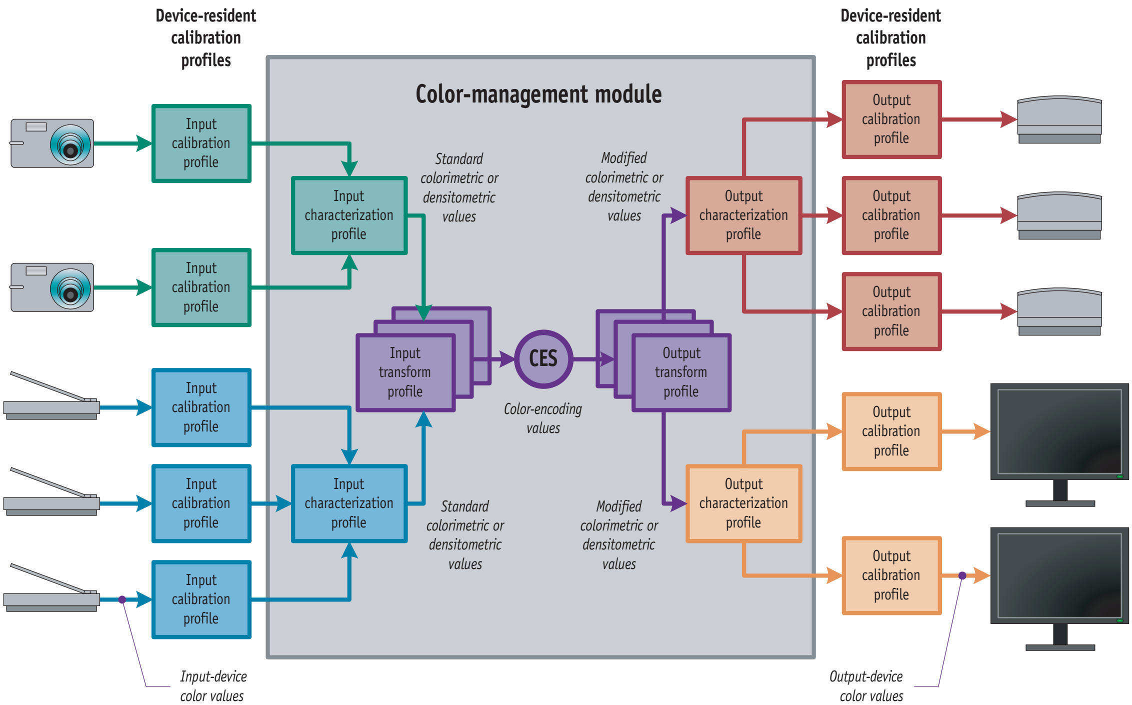


Figure 23.11c In this split-profile implementation, the calibration profiles reside in the input/output devices. This arrangement is recommended for systems having many input and/or output devices of the same basic type.

all output profiles would have to be replaced in the combined-profile approach previously shown in Figure 23.7. Similarly, in the advanced CMM approach shown in Figure 23.8, all CMMs would have to be updated, and in the advanced application approach shown in Figure 23.9, all applications would have to be updated.

Figure 23.12 illustrates one way in which the implementation concepts that have been discussed can be extended when necessary to meet the objectives of more complex systems. As the figure indicates, there are two CESs: one is designed specifically to meet the input requirements, and the other is designed specifically to meet the output requirements. This is the approach first described in the digital cinema example of Chapter 21. Here, it is shown in greater detail and includes the characterization and calibration operations required for practical implementation.

Current implementations

At the time of this writing, many of the ideas that have been discussed here have been put into practice. In particular, the concept of profiles and color-interchange spaces based on colorimetric values and associated sets of reference viewing conditions has been well received.

There are, however, two aspects of some current implementations and specifications that have prevented the adoption of the Unified Paradigm. The first is that some system CESs use media-relative colorimetry. As described earlier, the use of media-relative colorimetry does not produce a proper brightness adjustment of image data. As a result, the interpretation of CES values alone becomes ambiguous. Although it is possible to get around this problem, doing so requires the use of additional information to define the various media

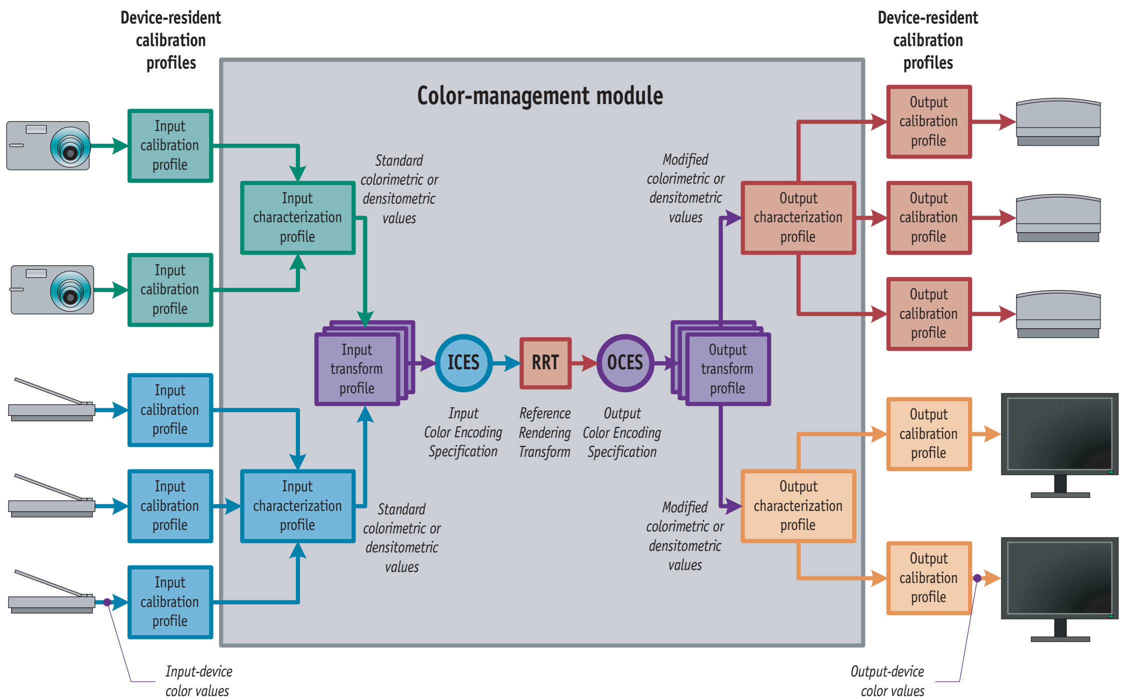


Figure 23.12 In this figure, the implementation concepts discussed in this chapter have been extended to meet the objectives of a more complex system. The system uses two CESs together with the characterization and calibration operations required for practical implementation.

white references used for encoding. Moreover, it results in a proliferation of different (and incompatible) image representations. We believe the overall color-management environment would be more robust if the unnecessary problems introduced by the use of different media-specific white references were eliminated. This could be done by using color-appearance encoding that is based on brightness-adapted colorimetric values, as described in Chapter 19.

The second problem is that some current specifications define the maximum relative luminance value encodable in the CES to be equivalent to a CIELAB L^* value of 100. As a result, the full dynamic ranges of current high-quality imaging devices and media are not fully preserved in the encoding. In our opinion, encoding specifications being developed today should at least support the capabilities of current media and devices. Moreover, the demands of new imaging devices and media having capabilities greater than those of current products should be anticipated and accommodated in the specifications. This requires the capability of representing extensive overall luminance dynamic ranges and maximum relative luminance values equivalent to CIELAB L^* values considerably greater than 100. These capabilities are of particular importance for input CESs, because it is at that stage of an imaging system where image color adjustments most likely are performed.

Discussion

A number of different approaches that can be used to implement the Unified Paradigm have been described in this chapter. It is important to emphasize that regardless of the approach used, the color transformations that are required to produce a given output image from a given input image do not fundamentally change when implemented using any of these different approaches. The approaches differ principally in whether the signal processing transformations are performed entirely by profiles alone, by color-management modules alone, by applications alone, or by various combinations of the three. The approaches also may differ in where the transforms reside.

It is especially important to emphasize that while different approaches can be used, any successful implementation of the Unified Paradigm must be based on a meaningful representation of color throughout the color-management environment. As was shown earlier, when the environment is defined to be truly global and completely unrestricted, that representation must unambiguously describe color itself.

Summary of key issues

- Implementation of a unified color-management environment can be facilitated by an industry-wide adoption of appropriately designed interchange standards and a common overall architecture.
- Basic architectures similar to those currently used within the industry can be used to implement the Unified Paradigm.
- Although various implementations may be employed, the color transformations that are required to produce a given output image from a given input image do not fundamentally change.
- Most importantly, the successful implementation of the Unified Paradigm must be based on a meaningful representation of color throughout the color-management environment. When the environment is defined to be global and unrestricted, that representation must unambiguously describe color itself.

24

Closing Thoughts and Conclusions

In the first edition of this book, published in January 1998, we offered the following assessment of the state of color management in the color-imaging industry:

> At the time of this writing, there is not yet an agreed-upon method for interchanging color images among different types of imaging systems. However, there has been a great deal of progress in that direction. Also encouraging is the fact that many of the technical problems discussed in this book are becoming more generally acknowledged and understood.
>
> In particular, there is growing acceptance of the two principal issues that were addressed: first, that digital color encoding involves much more than standard colorimetry; and second, that there is more to successful color interchange than the use of standard file formats and data metrics. There also is a much greater understanding of the need for appearance-based colorimetry, and the need to provide various input-encoding and output-rendering options is gradually becoming more recognized.
>
> Unfortunately, many problems remain. There still are misconceptions regarding the roles of color encoding specifications and color interchange specifications in color communication. There is confusion regarding the concept of encoding reference viewing conditions and the distinction between encoding reference conditions and actual input/output conditions. There also is confusion regarding specific concepts related to reference viewing conditions, such as the distinction between a reference illuminant and a reference adaptive white, the difference between media-relative colorimetry and brightness-adapted colorimetry, and the function of chromatic adaptation transforms. We sincerely hope that our book has helped to clarify these issues.
>
> Of some concern to us is that encoding based on color appearance will be misunderstood as being sufficient to provide a complete solution to color encoding, just as encoding based on standard colorimetry has often been misunderstood to do so. As discussed in the text, appearance-based color encoding provides a means for integrating other color-encoding methods, but it is not a substitute for those methods. Encoding methods that can extract original-scene colorimetry from reproduced images, that can render images from negatives and digital still cameras, and that can re-render reproduced images in various ways are required in order to provide the complete array of input-interpretation options defined by the Unified Paradigm. Again, we hope that our book has helped to explain such issues.

It has now been more than a decade since those words were written, yet our assessment of the state of color management today is little different. Much to our disappointment, the same misconceptions that have hindered progress for so long are still prevalent. As a result, while numerous color-management products and systems have been introduced during the intervening years, most have failed and quickly disappeared because they were based on

Digital Color Management: Encoding Solutions, Second Edition Edward J. Giorgianni and Thomas E. Madden
 ISBN: 978-0-470-51244-9

unsound understandings of color science and color imaging.

There have been some successes, but for the most part they have been limited to applications that have been narrow in their focus. For example, the sRGB color space has worked very well when used as intended, i.e., to represent digital code values for images originating on, or intended for display on, typical computer monitors or similar electronic displays. While the original developers of that encoding specification fully understood its purpose, others later tried to promote sRGB as a "universal" means of color encoding and interchanging images of all types. Predictably, those who attempted to use sRGB for such purposes soon encountered its inherent limitations. Again, the design of the sRGB specification was entirely appropriate for its intended purpose; the problem was its subsequent promotion for unintended and inappropriate uses. Numerous alternative encoding specifications similar to sRGB but with various extensions, primary sets, metadata additions, and other modifications followed later. Although some of these specifications have been useful for particular applications, none has been capable of supporting a truly comprehensive system for color management and image interchange, as described in the Unified Paradigm.

The color-management system most widely promoted within the industry during the intervening years has been the one developed by the International Color Consortium (ICC). The ICC is an industry group that was originally formed to define and administer a platform-independent format standard for profiles and to provide a means for registering associated profile tag signatures and descriptions. That original effort was quite successful and has been a significant contribution to the industry.

The ICC committee also developed a color-management system based on its profile specifications. At the beginning of that project, the stated objectives seemed to be comprehensive and consistent with our concept of a Unified Paradigm system. Whether those objectives ultimately were achieved or not is a matter of some controversy. Some ICC proponents consider the Unified Paradigm and its associated appearance-based color-encoding method to be fully embodied in the system produced by the ICC. Others have concluded that the inherent functionality of ICC color management is that of a Type *B* paradigm system and that the ICC color encoding specification (referred to as a *profile connection space* or PCS) is limited by its use of a reflection-print reference medium and media-relative colorimetry.

In response to these and other concerns, subsequent additions and revisions were made by the ICC in the form of various "rendering-intent" options, extensions, and other modifications. As a result, the system is now more versatile. However, it is also more complex and almost entirely dependent on the use of tagged information. As discussed earlier in this book, the use of tags alone does not provide a direct solution to input incompatibilities and other difficult color-management problems. Instead, tags provide information that *subsequent* applications may use in order to deal with difficulties not already addressed in a color encoding specification itself.

In order to avoid the apparent complexity of the ICC system, many practitioners have developed ways to take advantage of its standardized profile format while bypassing any ICC color management per se. For example, in applications dealing with images other than reflection prints (or softcopy images intended to look like reflection prints), practitioners may create their own PCSs that are not subject to limitations that might be imposed by a PCS based on a reflection-print reference medium and media-relative colorimetry. These unofficial PCSs are, of course, not directly compatible with the ICC PCS or with each other.

Alternatively, many practitioners use the ICC profile format and other aspects of the system to form device-link profiles that directly convert input-device code values to output-device code values. This long-established approach can be both efficient and effective, especially in applications where large numbers of images are to be processed through an identical chain of signal processing operations. However, the use of such profiles does not generate compatible intermediary image data that might be required for image compositing, manipulation, and archiving or for image interchange among different types of systems. The use of device-link profiles is supported in the ICC system, and such profiles can be formed by the concatenation of ICC input and ICC output profiles. However, if the connection between the two profiles is the ICC PCS, any limitations of the PCS encoding become imposed on the resulting concatenated profile. In some situations, important image information can be compromised. To avoid any such problems, many practitioners again have created their own unofficial PCSs specifically for use in developing and concatenating device-link profiles.

Whether such measures to circumvent ICC color-management methods and similar methods used in many other color-management products are fully warranted or not is a matter of opinion. What is certain is that there is a perceived need for these ad hoc solutions, and they are in widespread use. As a result, it would be difficult to conclude that the current state of the industry is significantly different from what it was in the past. For the most part, the industry still is effectively composed of numerous different types of systems operating more or less autonomously. Arguably, the overall situation may have worsened over the last decade because the industry's exponential growth during that period has led to the creation of even greater numbers of independent and often proprietary approaches that attempt to avoid real or perceived issues with available color encoding specifications and color-management systems. As a consequence, it now appears less likely that the single unified color-management environment—at least as we originally envisioned it—can be realized.

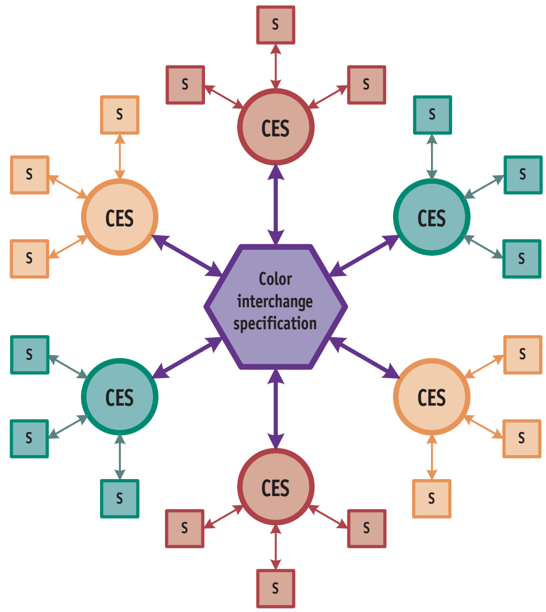


Figure 24.1 In this original vision of a unified color-management environment, related systems share common color encoding specifications (CESs), and all systems communicate through a single color interchange specification (CIS). Note that, in this figure and those that follow, arrows imply both the transfer of image information and the application of transforms appropriate for that transfer.

In that original vision, discussed in Chapter 22 and illustrated here in Figure 24.1, the interchange of color information among different groups was to have been accomplished using a standardized color interchange specification (CIS). To date, no CIS capable of fully supporting that vision has been made available to the industry, and the timeframe in which a single industry-wide standard for color interchange could be introduced likely has passed. Nevertheless, it may still be possible to adopt alternative implementations of the Unified Paradigm concept.

In one such alternative, various associated systems again would be grouped together, and different groups again would use CESs most appropriate for their specific needs. However, the method of color interchange would be fundamentally different. In this approach, shown in Figure 24.2, group-to-group interchange would be implemented by transforms that link specific pairs of CESs. Such linkages can be successful if, consistent with the color-encoding method of the Unified Paradigm, each CES explicitly or implicitly includes a complete description of its colorimetry, data metric, and associated viewing conditions. As discussed in Chapter 22, such descriptions would allow appropriate interchange transforms to be developed such that the transfer of color information from the CES of one system to that of another is meaningful and unambiguous.

Although this CES-to-CES interchange alternative is technically feasible, it is not especially desirable and would be difficult to implement due to the sheer number of alternative CESs (including

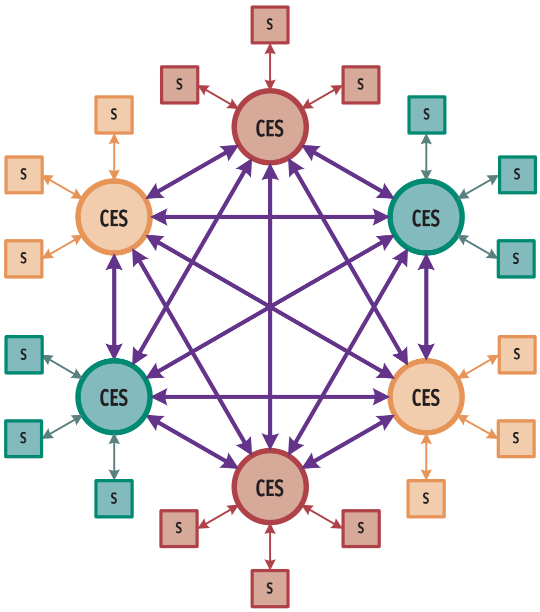


Figure 24.2 In this alternative implementation of a unified color-management environment, related systems again share common CESs, but systems communicate through numerous direct CES-to-CES interchange transformations.

numerous unofficial ICC PCS variations) that currently exist. The problem is similar to that discussed in Chapter 9 regarding multiple input/output systems, where the number of required connections can become prohibitive. Linking all combinations of just 10 different CESs, for example, would require 100 unique interchange transforms.

Figure 24.3a illustrates a better alternative that would be less disruptive to existing systems and thus may be more feasible to adopt at this time. As the figure shows, color interchange is implemented in this scheme using just three CISs, each associated with a different basic type of encoding reference viewing environment. One CIS is based on a set of conditions consistent with those of unrendered (scene-based) images. A second CIS is based on encoding reference viewing conditions associated with rendered images intended to be viewed in dark-surround environments (e.g., motion pictures, projection slides, home theater, etc.). The third CIS is based on encoding reference viewing conditions associated with rendered images intended to be viewed in average-surround environments (e.g., reflection prints, some monitor displays, etc.). The use of different CISs associated with different image states and encoding reference viewing environments would simplify the interchange of images among systems having CESs that are fundamentally similar but that differ in their particulars.

To be consistent with the objectives of the Unified Paradigm, interchange among dissimilar system groups also must be supported. This could be accomplished by the addition of transforms connecting the three CISs (Figure 24.3b). Adoption of this implementation would be most likely if the CISs were based on appropriate available standards. If, for example, the digital cinema system ultimately recommended by the Academy of Motion Picture Arts and Sciences is equivalent to that described in Chapter 21, the Academy ICES could serve as the scene-based CIS and the Academy OCES could serve as the rendered image dark-surround CIS. At the time of this writing, there is no available rendered image CIS that is entirely consistent with the requirements we have described for an average-surround CIS. However, developing an appropriate specification would be relatively straightforward. The example CIS described in Chapter 22, although originally intended for general interchange in the Unified Paradigm, would also be well suited for this more specific application. An implementation based on these particular interchange specifications is shown in Figures 24.3c.

The two-way arrows connecting the three CISs in Figure 24.3c indicate that interchange among these specifications is unrestricted and bidirectional. However, it is important to recognize that interchange from the average-surround CIS *to* the ICES or OCES may not always be appropriate. Some restrictions could be due to the average-surround CIS itself, although if the principles described in Chapter 22 are followed that is not inevitable. Other limitations, however, may be unavoidable due to the nature of certain CESs supported by that CIS. For example, image information defined in terms of a typical CMYK-based graphic arts CES or other reflection-print-based CES, transformed to the average-surround

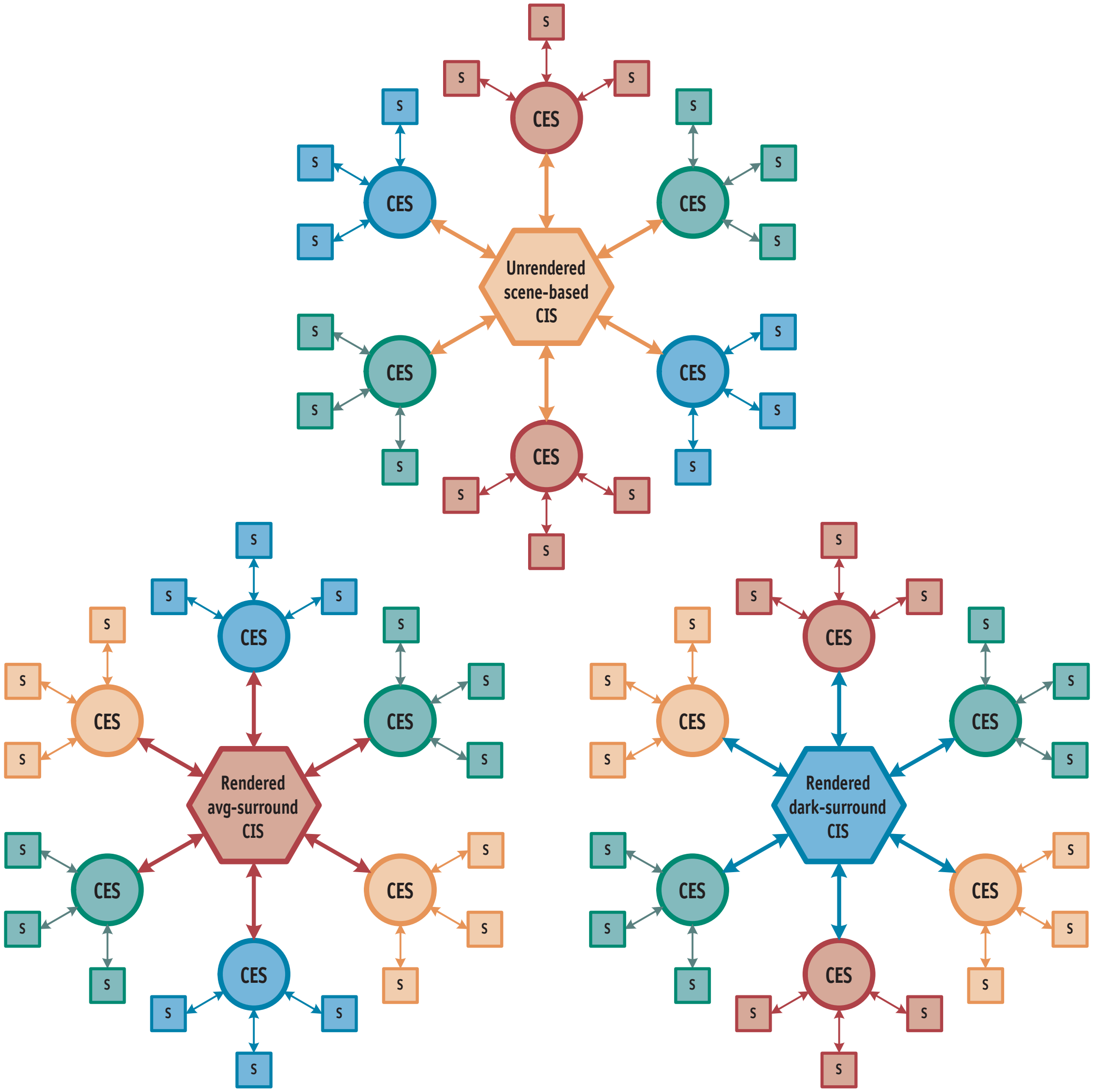


Figure 24.3a In this alternative implementation of a unified color-management environment, related systems once again share common CESs. In addition, related groups share common CISs associated with the encoding reference viewing environments of the CESs involved.

CIS, and then transformed to ICES, likely would be inadequate for most applications directly associated with that extensive scene-based specification. However, such interchange should not be disallowed completely. It might be acceptable, for example, to transfer an image encoded in terms of a reflection-print-based CES to ICES for use as a backdrop within a digital cinema composite image.

* * *

Because we are older (definitely) and a bit wiser (perhaps), we now better appreciate that major

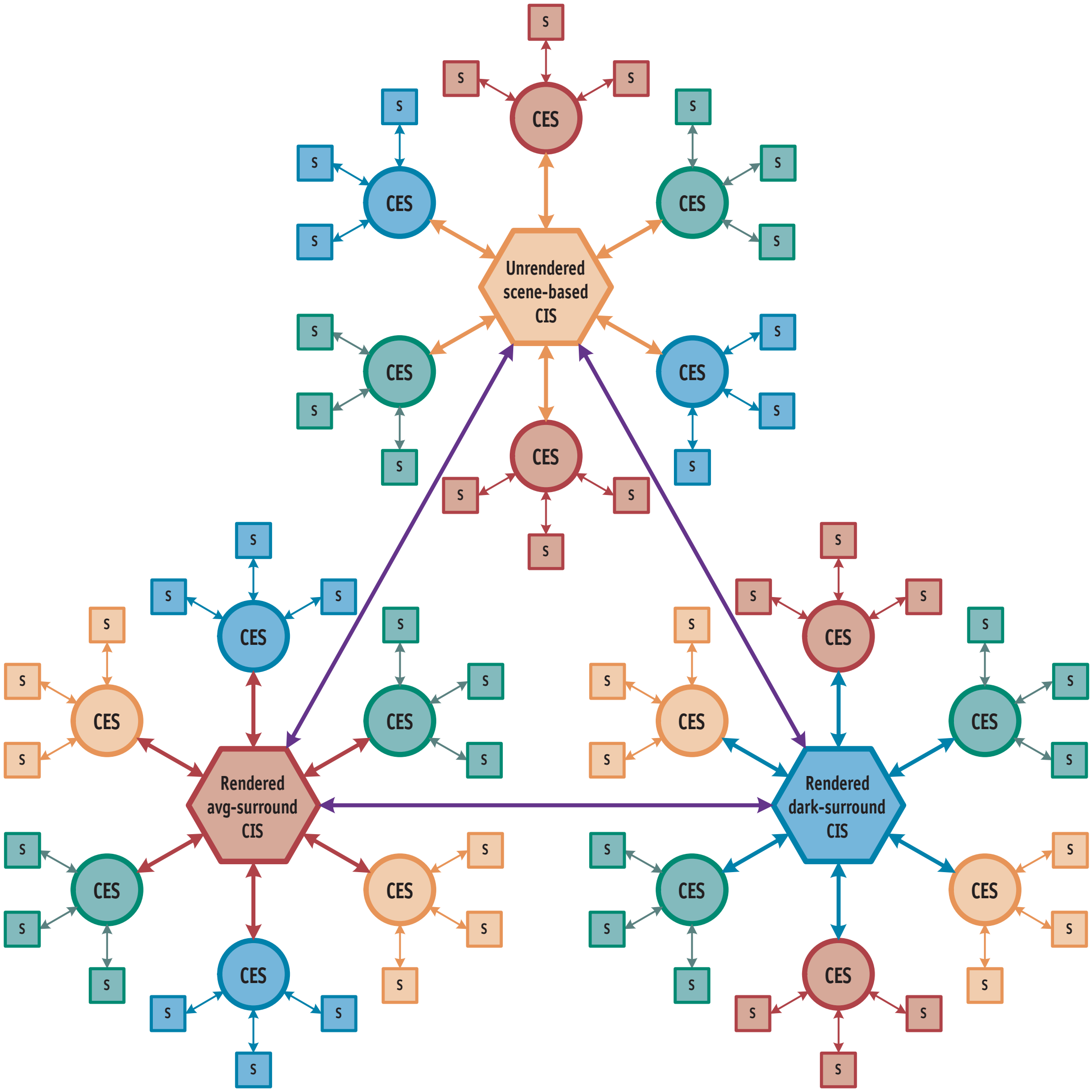


Figure 24.3b Interchange among the dissimilar groups of systems operating in the unified color-management environment of Figure 24.3a can be provided by the addition of transforms connecting the three CISs.

changes such as those we proposed in the first edition of this book can take longer than expected to be understood, accepted, and implemented. So despite our misgivings about how events have proceeded over the years, we have not lost our enthusiasm for this subject. Nor have we lost our optimism or our hope that someday the promise of the Unified Paradigm will be fully realized—perhaps based on one of the implementations just described. With such positive thoughts in mind, we will close with the same words we used to conclude our first edition:

> Despite these concerns, we are optimistic that our ultimate goal of a unified color-management environment for the color-imaging industry someday will be reached. We envision a global environment encompassing a complete hierarchy of

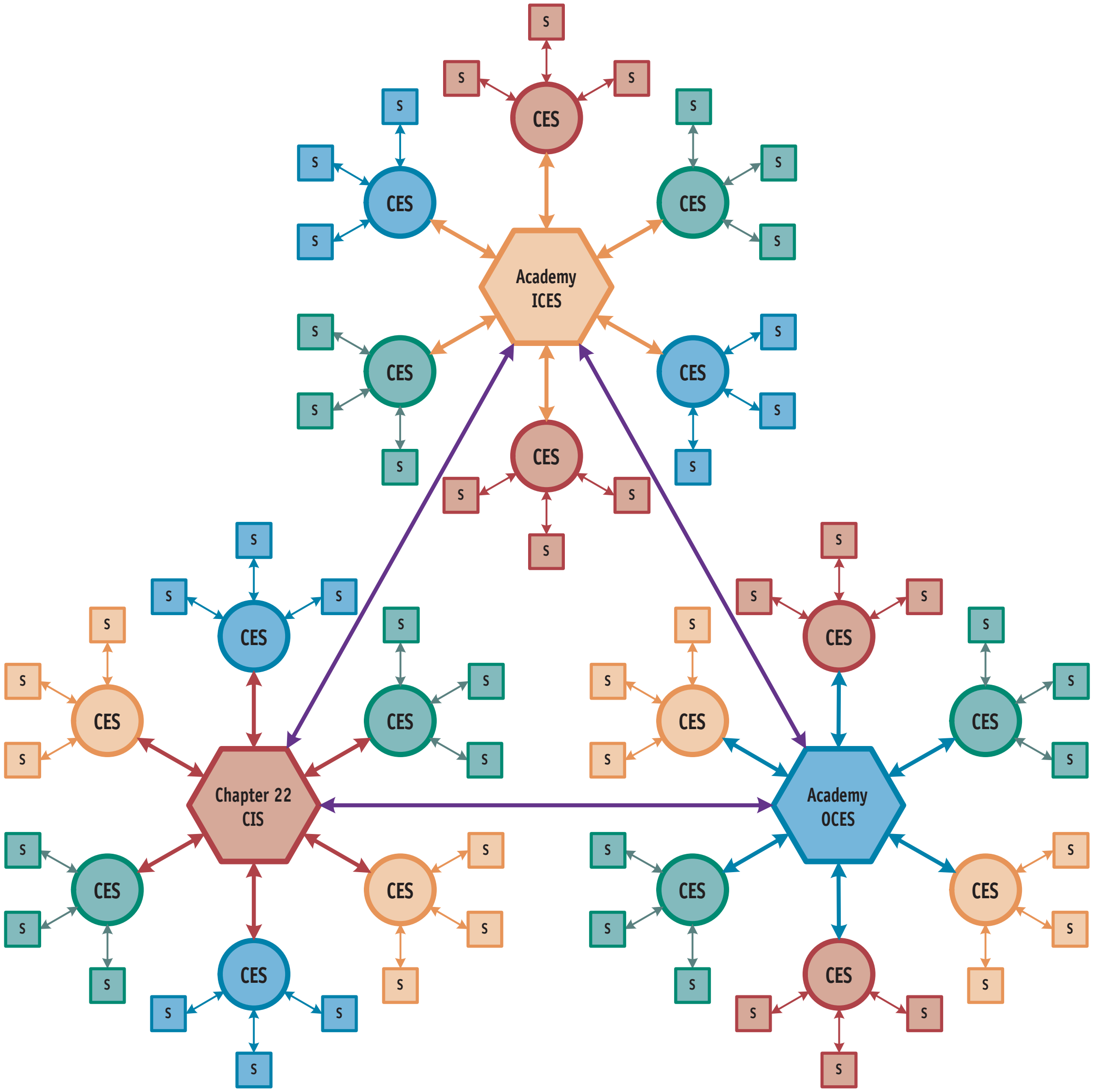


Figure 24.3c One possible realization of the unified color-managed environment of Figures 24.3a and 24.3b. Here, a digital cinema ICES serves as the scene-based CIS, a digital cinema OCES serves as the dark-surround CIS, and the CIS described in Chapter 22 serves as the average-surround CIS.

imaging systems and applications, from the simplest to the most sophisticated, all operating according to a single underlying paradigm. In this environment, the unifying force will be the unrestricted and unambiguous representation of color.

When this goal is realized, the full potential of digital color imaging will be achievable; and when that happens, imaging applications that we can only imagine today will be accepted as commonplace.

PART V

Appendices

Digital Color Management: Encoding Solutions, Second Edition Edward J. Giorgianni and Thomas E. Madden
 ISBN: 978-0-470-51244-9

A

Colorimetry

References made throughout this book to *standard colorimetry, or to standard colorimetric values*, refer to colorimetric values determined according to CIE (Commission Internationale de l'Éclairage) recommended practices. All standard colorimetric values shown have been determined using the color-matching functions for the CIE 1931 Standard Colorimetric Observer (Figure A.1), whose color-matching characteristics are representative of those of the human population having normal color vision. The CIE 1931 Standard Colorimetric Observer is often referred to as the *2° Observer*, because test fields subtending a viewing angle of 2° were used in the judging experiments from which the color-matching functions were derived.

The CIE also has defined the 1964 Supplementary Standard Colorimetric Observer (Figure A.2), often referred to as the *10° Observer*. The color-matching functions of the 10° Observer are used for colorimetric measurements and calculations related to relatively large areas of color. The color-matching functions of the 2° Observer are used for most colorimetric measurements and calculations related to pictorial and graphics imaging, where individual areas of color generally subtend relatively small viewing angles.

The 2° Observer color-matching functions are used in the calculation of CIE tristimulus values X, Y, and Z, which quantify the trichromatic characteristics of color stimuli. The X, Y, and Z tristimulus values for a given object (characterized by its spectral reflectance or transmittance) that is illuminated by a light source (characterized by its spectral power distribution) can be calculated for the 2° Observer (characterized by the appropriate set of CIE color-matching functions) by summing the products of these distributions over the visible wavelength (λ) range (usually from 380 to 780 nm, at 5 nm intervals). The calculation of X, Y, and Z values is shown in Equations (A.1), and the basic procedure is diagrammed in Figure A.3:

$$\begin{aligned} X &= k \sum_{\lambda=380}^{780} S(\lambda)\, R(\lambda)\, \bar{x}(\lambda) \\ Y &= k \sum_{\lambda=380}^{780} S(\lambda)\, R(\lambda)\, \bar{y}(\lambda) \\ Z &= k \sum_{\lambda=380}^{780} S(\lambda)\, R(\lambda)\, \bar{z}(\lambda) \end{aligned} \tag{A.1}$$

where X, Y, and Z are the CIE tristimulus values; $S(\lambda)$ is the spectral power distribution of the light source; $R(\lambda)$ is the spectral reflectance or transmittance of the object; $\bar{x}(\lambda)$, $\bar{y}(\lambda)$, and $\bar{z}(\lambda)$ are the color-matching functions of the 2° Observer; and k is a normalizing factor. By convention, k generally is determined such that $Y = 100$ when the object is *a perfect white*. A perfect white is an ideal, nonfluorescent, isotropic diffuser with a reflectance

Digital Color Management: Encoding Solutions, Second Edition Edward J. Giorgianni and Thomas E. Madden
 ISBN: 978-0-470-51244-9

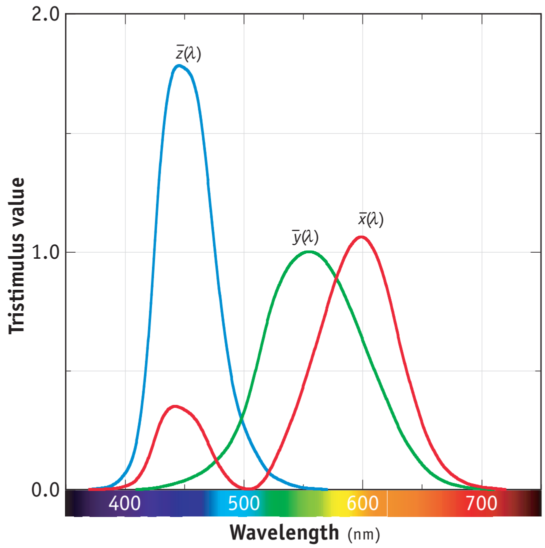


Figure A.1 Color-matching functions for the CIE 1931 Standard Colorimetric Observer.

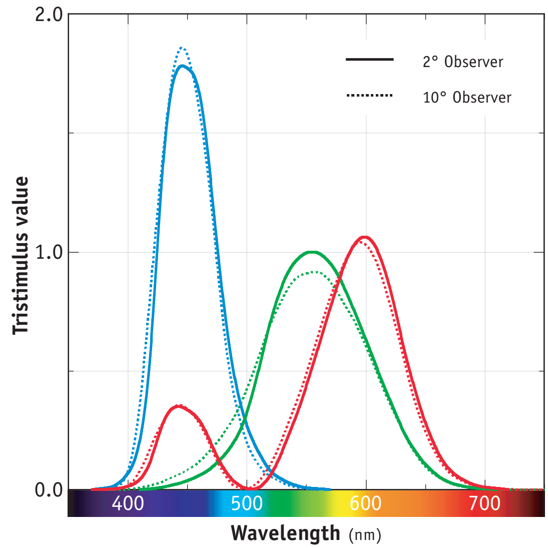


Figure A.2 Color-matching functions defined for the CIE 1964 Supplementary Standard Colorimetric Observer (10°) and the CIE 1931 Standard Colorimetric Observer (2°).

(or transmittance) equal to unity throughout the visible spectrum.

Chromaticity coordinates x, y, and z are derived from the tristimulus values as follows:

$$x = \frac{X}{X+Y+Z}$$
$$y = \frac{Y}{X+Y+Z} \tag{A.2}$$
$$z = \frac{Z}{X+Y+Z}$$

The CIE also has recommended the use of other coordinate systems, derived from *XYZ*, in which visual differences among colors are more uniformly represented. These systems include the CIE 1976 u', v' uniform-chromaticity-scale diagram and the CIE 1976 $L^*a^*b^*$ (CIELAB) color space.

In this book, u', v' diagrams (Figure A.4) are used for describing the chromaticities of color primaries, the chromaticity boundaries of color gamuts, and the results of applying chromatic adaptation transformations. The chromaticity values u' and

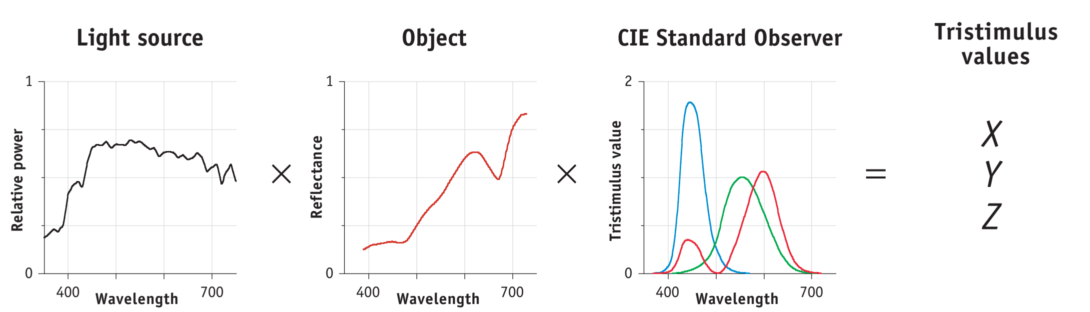


Figure A.3 Calculation of CIE XYZ tristimulus values.

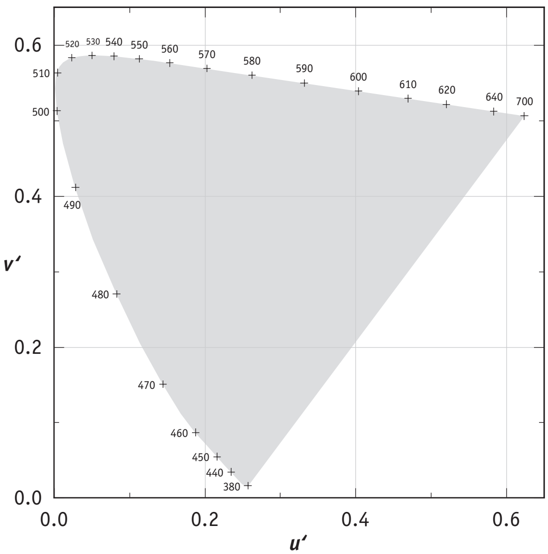


Figure A.4 CIE u', v' chromaticity diagram.

v' are computed from XYZ tristimulus values as follows:

$$u' = \frac{4X}{X + 15Y + 3Z}$$
$$v' = \frac{9Y}{X + 15Y + 3Z} \qquad \text{(A.3)}$$

Also in this book, CIELAB values are used for making comparisons of colorimetric color reproductions and for describing device and media color gamuts in three dimensions. CIELAB L*, a^*, and b^* values are computed from the tristimulus values X, Y, and Z of a stimulus, and the tristimulus values X_n, Y_n, and Z_n of the associated reference white, as follows:

$$L^* = 116\left(\frac{Y}{Y_n}\right)^{\frac{1}{3}} - 16 \quad \text{for } \frac{Y}{Y_n} \geq 0.008856$$
$$L^* = 903.3\left(\frac{Y}{Y_n}\right) \quad \text{for } \frac{Y}{Y_n} < 0.008856$$

and

$$a^* = 500\left[f\left(\frac{X}{X_n}\right) - f\left(\frac{Y}{Y_n}\right)\right]$$
$$b^* = 200\left[f\left(\frac{Y}{Y_n}\right) - f\left(\frac{Z}{Z_n}\right)\right] \qquad \text{(A.4)}$$

where

$$f\left(\frac{X}{X_n}\right) = \left(\frac{X}{X_n}\right)^{\frac{1}{3}} \quad \text{for } \frac{X}{X_n} \geq 0.008856$$
$$f\left(\frac{X}{X_n}\right) = 7.787\left(\frac{X}{X_n}\right) + \frac{16}{116} \quad \text{for } \frac{X}{X_n} < 0.008856$$
$$f\left(\frac{Y}{Y_n}\right) = \left(\frac{Y}{Y_n}\right)^{\frac{1}{3}} \quad \text{for } \frac{Y}{Y_n} \geq 0.008856$$
$$f\left(\frac{Y}{Y_n}\right) = 7.787\left(\frac{Y}{Y_n}\right) + \frac{16}{116} \quad \text{for } \frac{Y}{Y_n} < 0.008856$$
$$f\left(\frac{Z}{Z_n}\right) = \left(\frac{Z}{Z_n}\right)^{\frac{1}{3}} \quad \text{for } \frac{Z}{Z_n} \geq 0.008856$$
$$f\left(\frac{Z}{Z_n}\right) = 7.787\left(\frac{Z}{Z_n}\right) + \frac{16}{116} \quad \text{for } \frac{Z}{Z_n} < 0.008856$$

Throughout this book, colorimetric comparisons of original and reproduced color stimuli are illustrated in terms of vector arrows on CIELAB a^*, b^* diagrams, such as that shown in Figure A.5. In that diagram, the a^*, b^* coordinates of the original

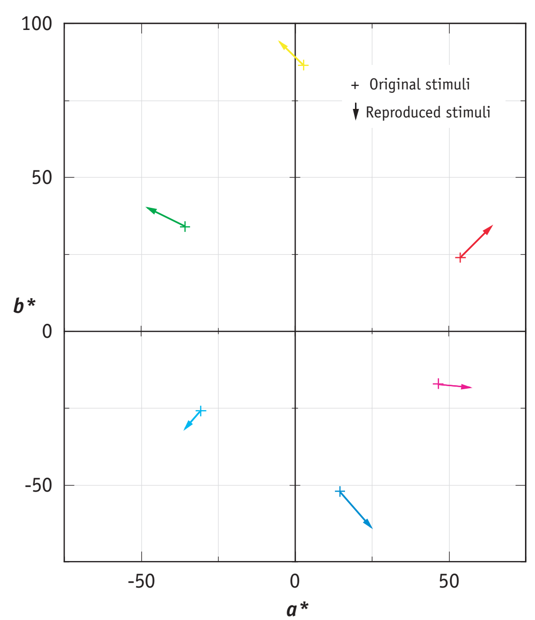


Figure A.5 In CIELAB a^*, b^* diagrams such as this, the length of a vector arrow connecting two stimuli indicates their chromatic difference.

color stimuli are represented by the + marks at the tails of the vector arrows. The heads of the arrows represent the coordinates of the corresponding reproduced color stimuli. The lengths of the connecting vectors indicate the magnitudes of the *chromatic* (hue and chroma) differences between the original stimuli and their reproductions. Note, however, that these vectors alone do not completely describe the *colorimetric* differences of the stimuli. Differences in L^* are part of the total colorimetric difference, and L^* differences are not shown on a CIELAB a^*, b^* diagram.

B
Densitometry

Throughout this book, the word *density* refers specifically to optical density. Optical density measurements and density-measuring devices (densitometers) are important in color imaging for several reasons: most forms of color encoding are based directly or indirectly on density readings; most input and output devices are calibrated using density measurements; and most reflection and transmission scanners essentially are densitometers.

Figure B.1 illustrates the basic concept of optical density measurement. A densitometer consists of a light source, a means for inserting a sample to be measured, a photodetector, an amplifier, and some type of analog or digital density-value indicator. Density readings are made by measuring the amount of light with and without the sample in place.

For transmissive samples, the transmission density is determined from the transmittance factor of the sample. The transmittance factor, T, is the ratio of the amount of light transmitted, I_t, measured with the sample in place, to the amount of incident light, I_i, measured without the sample (Figure B.2). The transmission density, D_t, of the sample is the negative logarithm of its transmittance factor:

$$T = \frac{I_t}{I_i} \tag{B.1a}$$

$$D_t = \log_{10}\left(\frac{1}{T}\right) \quad \text{or} \quad -\log_{10} T \tag{B.1b}$$

The following examples may help to illustrate the basic relationships of sample transmission characteristics, measured transmittance factor, and computed transmission density:

- If a sample were a perfect transmitter, the amount of incident and transmitted light would be equal, the measured transmittance factor would be 1.00, and the computed transmission density of the sample would be zero.
- If a sample were perfectly opaque, the amount of transmitted light would be zero, the measured transmittance factor would be zero, and the computed transmission density would be infinite.
- If a sample transmitted one-quarter of the incident light, the measured transmittance factor would be 0.25, and the computed transmission density would be $-\log_{10}(0.25)$ or 0.602.

Reflection samples are measured similarly (Figure B.3). The ratio of the amount of light reflected by a sample, I_r, relative to that reflected by a perfect diffuse reflector, I_{ref}, is the reflectance factor, R. Reflection density, D_r, is the negative logarithm of the reflectance factor:

$$R = \frac{I_r}{I_{\text{ref}}} \tag{B.2a}$$

$$D_r = \log_{10}\left(\frac{1}{R}\right) \text{ or } -\log_{10} R \tag{B.2b}$$

Digital Color Management: Encoding Solutions, Second Edition Edward J. Giorgianni and Thomas E. Madden
 ISBN: 978-0-470-51244-9

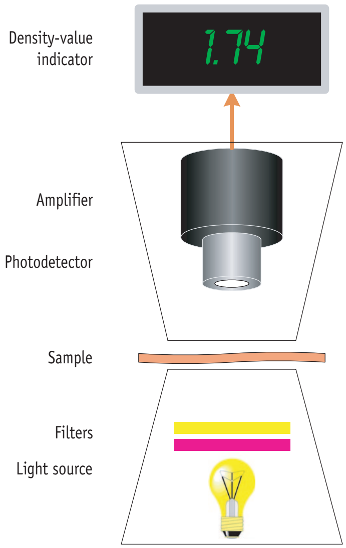


Figure B.1 Components of a basic optical densitometer.

The concept of density is quite straightforward for samples that are spectrally nonselective, i.e., samples that transmit (or reflect) light equally at all wavelengths of interest (Figure B.4). Since the transmittance or reflectance—and therefore the density—of

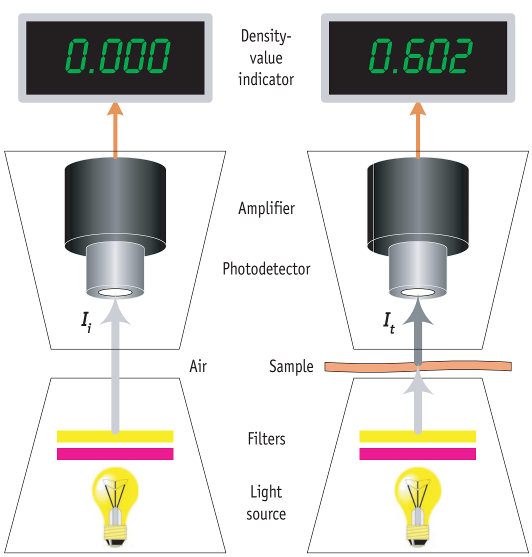


Figure B.2 Measurement of transmission density.

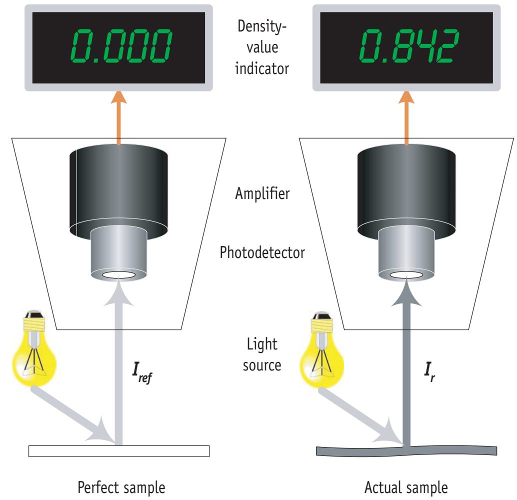


Figure B.3 Measurement of reflection density.

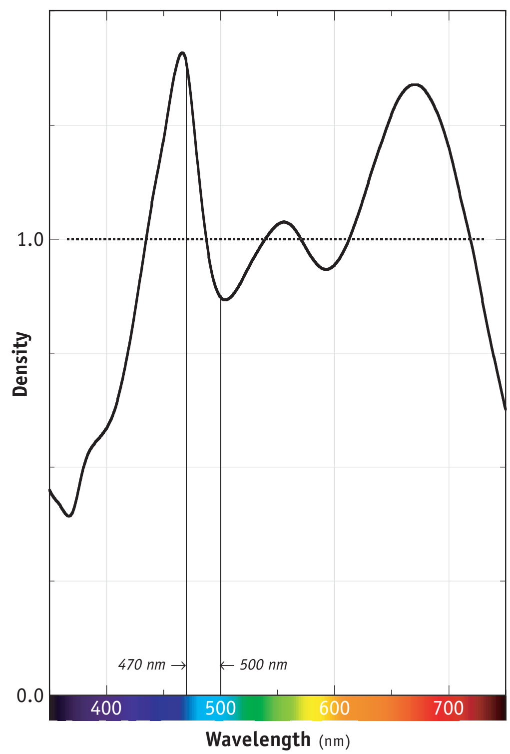


Figure B.4 This spectrally selective sample (solid line) has a density of 1.39 at 470 nm, but a density of only 0.87 at 500 nm. A spectrally nonselective sample, such as the one shown by the dotted line, has equal densities at all wavelengths of interest.

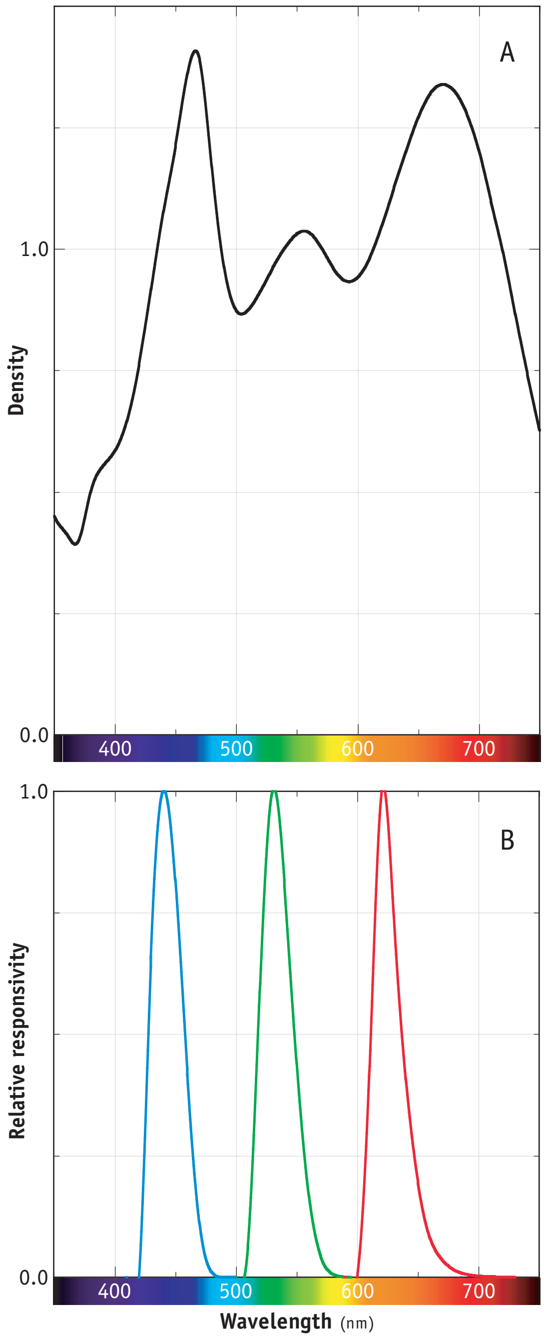


Figure B.5 A spectrally selective sample and a set of red, green, and blue spectral responsivities. The sample (A) has RGB density values of 1.10, 0.97, and 1.11 when measured by a densitometer having the RGB responsivities shown in B.

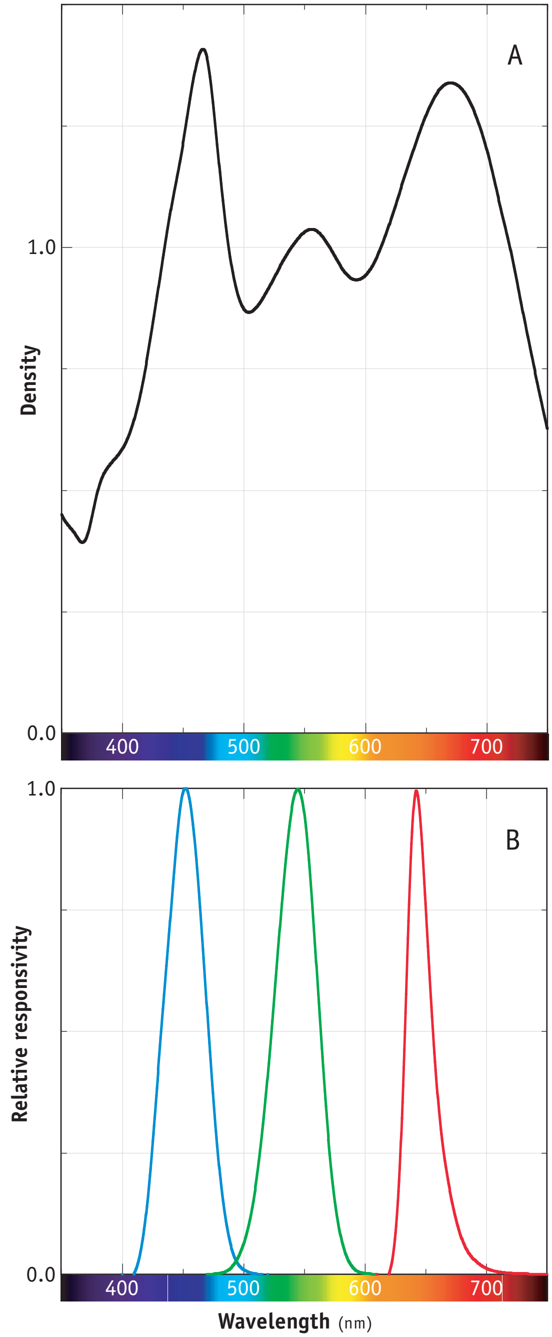


Figure B.6 A spectrally selective sample, identical to that shown in Figure B.5, and a different set of red, green, and blue spectral responsivities. The sample (A) has RGB density values of 1.24, 1.00, and 1.16 when measured by a densitometer having the RGB responsivities shown in B.

such samples is the same at every wavelength, it will not matter which wavelengths are used in their measurement.

In imaging applications, however, samples generally are made up of dyes, inks, or pigments that are spectrally selective, like the selective sample also shown in Figure B.4. When the density of a sample varies with wavelength, it no longer can be said that the sample has a particular density unless the spectral characteristics of the measurement also are specified. For example, the spectrally selective sample in Figure B.4 has a density of 1.39 if measured at 470 nm, but it has a density of only 0.87 if measured at 500 nm.

For color measurements, red, green, and blue densities generally are measured. But, again, the spectral characteristics of those measurements must be specified. For example, in Figure B.5, the sample has *RGB* densities of 1.10, 0.97, and 1.10 respectively when measured using a densitometer having the spectral responsivities shown. In Figure B.6, the same sample has RGB densities of 1.24, 1.00, and 1.16 respectively when measured using a densitometer having somewhat different spectral responsivities.

In this book, several different types of density measurements, based on different spectral responsivities, are used: visual densities; ISO Status A *and* Status M densities; scanner densities; and printing densities.

Visual densities, which are used throughout the book to express grayscale characteristics, are measured according to a responsivity equivalent to the CIE *XYZ* $\bar{y}(\lambda)$ function (Figure B.7). This measurement yields the CIE *XYZ* *Y* tristimulus value, and visual density, D_v, is computed from that value as follows:

$$D_v = -\log_{10}\left(\frac{Y}{100}\right) \tag{B.3}$$

Status A densities are measured using the red, green, and blue responsivities specified by the International Organization for Standardization (ISO) for Status A densitometers (Figure B.8). Status A densitometers are used in measuring photographic reflection prints and slides and other types of imaging media that are designed to be viewed directly. Note, however, that these responsivities are spectrally narrow. They were not designed to correspond to visual sensitivities; they were determined such that each channel primarily measures just one image-forming dye. The blue channel primarily measures the relative amount of yellow dye. Similarly, the green channel primarily measures the relative amount of magenta dye, and the red channel primarily measures the relative amount of cyan dye. This type of measurement is useful for monitoring media manufacturing operations and subsequent image-forming processes.

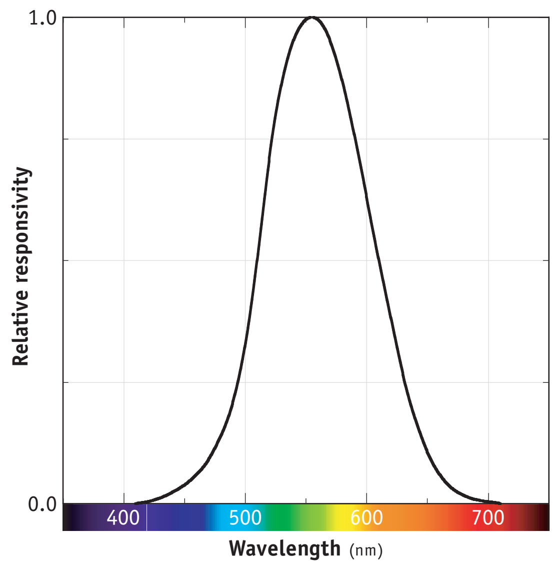


Figure B.7 The CIE *XYZ* $\bar{y}(\lambda)$ function, sometimes called the spectral luminous efficiency function, $V(\lambda)$.

Status M densities are measured using the red, green, and blue responsivities specified by the ISO for Status M densitometers (Figure B.9). Status M densitometers are used in measuring photographic negative media. Like those of Status A densitometers, the responsivities of Status M densitometers are spectrally narrow so that each channel primarily measures just one image-forming dye. Status M measurements are useful for monitoring negative-film manufacturing operations and subsequent chemical processing.

Scanner densities are densities measured by an input scanner having a particular set of *effective spectral responsivities*, *ESR* (λ). The effective responsivities are the product of the spectral sensitivity of the sensor and the spectral characteristics of any

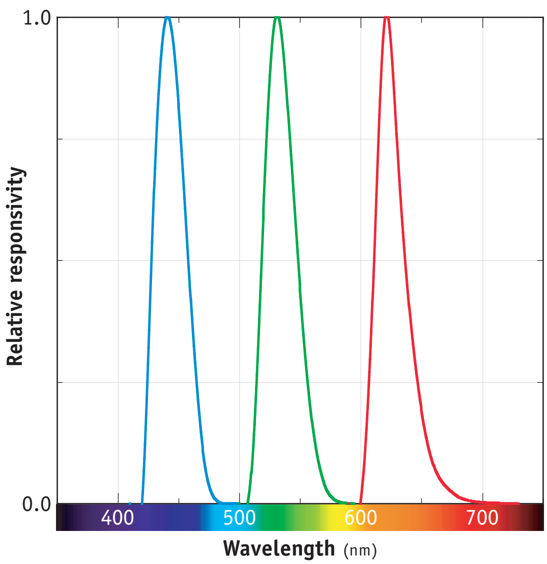


Figure B.8 Spectral responsivities for an ISO Status A densitometer.

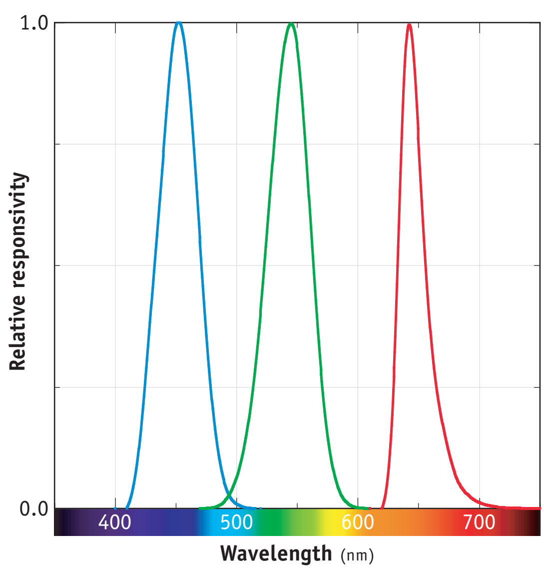


Figure B.9 Spectral responsivities for an ISO Status M densitometer.

components in the optical chain that contribute to the overall spectral responses. These components include the scanner light source, lenses, mirrors, and filters. For example, the effective responsivity for one of the color channels of a scanner would be

$$ESR\,(\lambda) = S\,(\lambda)\,M\,(\lambda)\,L\,(\lambda)\,F\,(\lambda)\,SS\,(\lambda) \qquad \text{(B.4)}$$

where $ESR(\lambda)$ is the effective responsivity of the channel, $S\,(\lambda)$ is the spectral power distribution of the scanner light source, $M\,(\lambda)$ is the spectral reflectance of a mirror, $L\,(\lambda)$ is the spectral transmittance of a lens, $F\,(\lambda)$ is the spectral transmittance of a filter (or a series of filters) in the optical path of the particular color channel, and $SS\,(\lambda)$ is the spectral sensitivity of the sensor.

Printing densities are the densities that would be measured by a real or computational densitometric device having effective spectral responsivities equivalent to those of a particular photographic print medium and a specified printer. These effective responsivities are the product of the spectral sensitivity of the print medium and the spectral characteristics of any components in the printer that contribute to the overall spectral responses, such as the printer light source, lenses, mirrors, and filters. For example, the effective red spectral responsivity would be

$$ESR_r\,(\lambda) = S\,(\lambda)\,M\,(\lambda)\,L\,(\lambda)\,F\,(\lambda)\,SS_r\,(\lambda) \qquad \text{(B.5)}$$

where $ESR_r\,(\lambda)$ is the red-channel effective responsivity, $S\,(\lambda)$ is the spectral power distribution of the printer light source, $M\,(\lambda)$ is the spectral reflectance of a mirror in the printer optics, $L\,(\lambda)$ is the spectral transmittance of the printer lens, $F\,(\lambda)$ is the spectral transmittance of a filter (or a series of filters) in the optical path of the particular color channel, and $SS_r\,(\lambda)$ is the red spectral sensitivity of the print medium.

The red transmittance factor T_r, for a sample having the spectral transmittance measured according to the red-channel effective spectral responsivity, can be calculated as follows:

$$T_r = k_r \sum_{\lambda} T\,(\lambda)\,ESR_r\,(\lambda) \qquad \text{(B.6)}$$

where k_r is a normalizing factor, determined such that $T_r = 1.00$ when the sample is a perfect transmitter. The red printing density, PD_r, of the sample

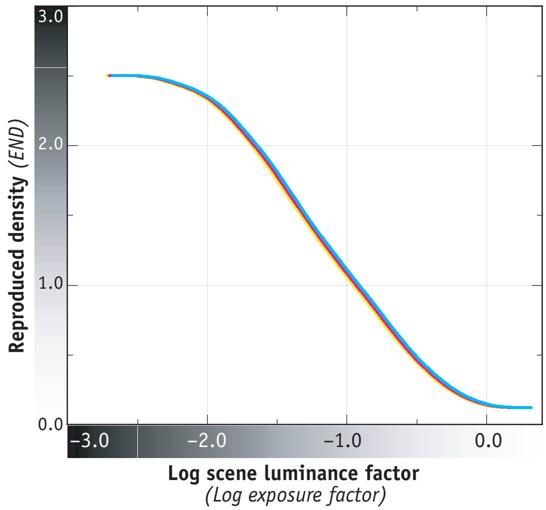


Figure B.10 The equal equivalent neutral density (END) values of the cyan, magenta, and yellow colorants of this grayscale indicate that the grayscale is colorimetrically neutral throughout its exposure range.

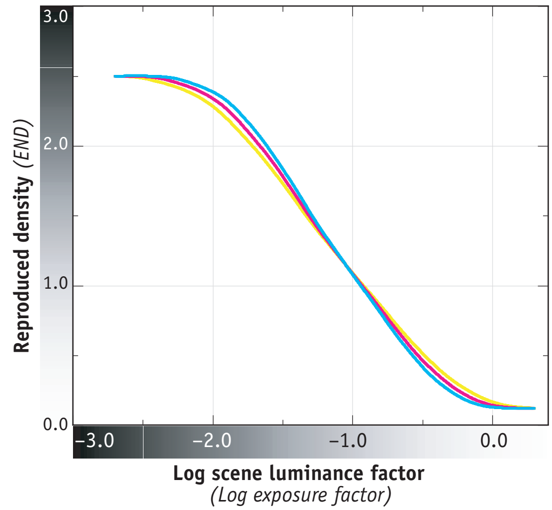


Figure B.11 The unequal equivalent neutral density (END) values of the cyan, magenta, and yellow colorants of this grayscale indicate that the grayscale is not colorimetrically neutral throughout its exposure range. The grayscale is colorimetrically cyan-blue at lower exposure levels and colorimetrically yellow-red at higher exposure levels.

then can be calculated as follows:

$$PD_r = -\log_{10}(T_r) \tag{B.7}$$

Green and blue effective responsivities, transmittance factors, and printing-density values can be calculated similarly.

Equivalent neutral densities (ENDs) are a measure of the relative amounts of the colorants of a medium. Throughout this book, grayscales or color scales plotted in terms of cyan, magenta, and yellow densities are expressed in terms of END values. The END value of a given amount of one colorant is the visual density that would result if the appropriate amounts of the other colorants were added such that a colorimetric neutral is formed, based on a specified viewing illuminant. In Figure B.10, for example, the cyan, magenta, and yellow END values are equal to each other throughout the grayscale characteristic curve. This means that at every exposure level, each corresponding dye amount has just the right amounts of the other two dyes to form a colorimetric neutral, which indicates that the grayscale is colorimetrically neutral throughout its exposure range. In Figure B.11, the unequal END values of the cyan, magenta, and yellow scales indicate that the grayscale is not colorimetrically neutral at every exposure level. At higher exposures, the comparatively high yellow and low cyan END values indicate that the grayscale is colorimetrically yellow-red in that region. At lower exposures, the comparatively high cyan and low yellow END values indicate that the grayscale is colorimetrically cyan-blue in that region. As discussed in Chapter 7, in some situations it can be preferable for a grayscale not to be colorimetrically neutral and thus to have unequal END values.

C

Photographic Media

Three principal types of conventional photographic media are referred to throughout this book: photographic *negative films*, such as those widely used in still and motion picture cameras; photographic *print media*, such as photographic papers and print films respectively used for making reflection and transmission prints from photographic negative films; and photographic *transparency films*, such as projection slide films and larger-format sheet films generally used with back illumination.

Photographic negative films and print media

Figure C.1 shows a simplified cross-section of a photographic negative film. The film has a blue-light-sensitive layer, a green-light-sensitive layer, and a red-light-sensitive layer coated on a transparent support. Because the green-light-sensitive and red-light-sensitive layers are also inherently sensitive to blue light, a yellow filter layer is coated above these layers to prevent blue light from reaching them. The yellow filter layer is made colorless during subsequent chemical processing of the film.

The light sensitivity of each layer is due to the silver halide grains that are dispersed within the layer. The grains of each layer are chemically treated during the manufacturing process in order to produce the appropriate spectral sensitivity for the particular layer. The resulting red, green, and blue spectral sensitivities for a typical photographic negative film are shown in Figure C.2. Also dispersed in each layer is an appropriate *coupler*. The function of the coupler will be explained shortly.

When an exposed photographic negative film is chemically processed, yellow, magenta, and cyan image dyes are formed in the blue-light-sensitive, green-light-sensitive, and red-light-sensitive layers, respectively. The nature of the chemical processing is such that a negative image results, i.e., the *maximum* amount of dye forms at the *maximum* exposure; the *minimum* amount of dye forms at the *minimum* exposure. Also during chemical processing, the yellow filter layer is made colorless.

Figure C.3(a) shows a series of color stimuli that are to be photographed by a photographic negative film. Figure C.3(b) shows how each layer of the film is affected by exposure to the stimuli. The red stimulus exposes primarily the red-light-sensitive (bottom) layer, leaving the other layers essentially unexposed. Similarly, the green stimulus exposes primarily the green-light-sensitive (middle) layer, and the blue stimulus exposes primarily the blue-light-sensitive (top) layer. The other stimuli represent mixtures of additive primaries, and thus they will expose more than one layer of the negative.

The cyan (green plus blue) stimulus exposes primarily the green-light-sensitive and blue-light-sensitive layers, leaving the red-light-sensitive layer

Digital Color Management: Encoding Solutions, Second Edition Edward J. Giorgianni and Thomas E. Madden

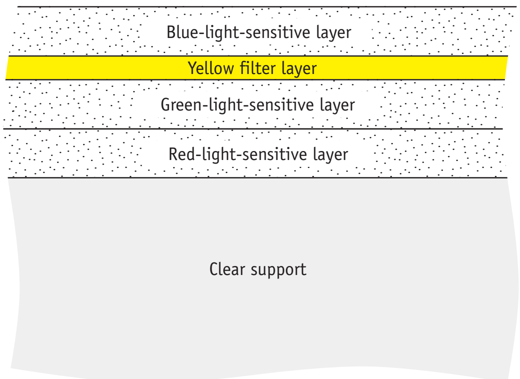


Figure C.1 Simplified cross-section of a color-negative film (prior to chemical processing).

essentially unexposed. Similarly, the magenta (red plus blue) stimulus exposes primarily the red-light-sensitive and blue-light-sensitive layers, leaving the green-light-sensitive layer essentially unexposed, and the yellow (red plus green) stimulus exposes primarily the red-light-sensitive and the green-light-sensitive layers, leaving the blue-light-sensitive layer

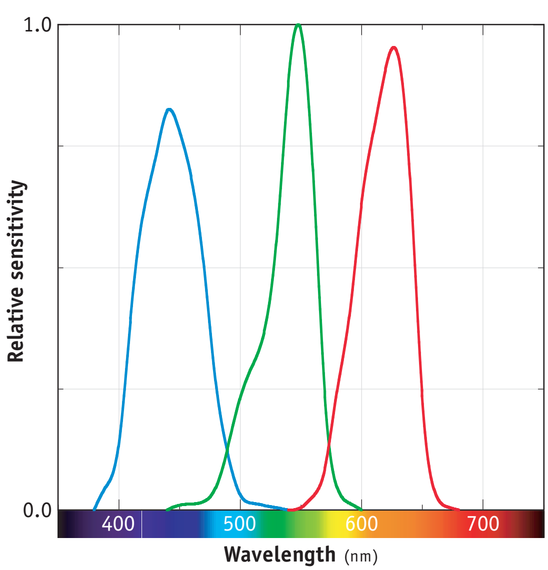


Figure C.2 Spectral sensitivities for a typical photographic negative film.

essentially unexposed. The white (red plus green plus blue) stimulus exposes all three layers, and a black stimulus leaves all layers essentially unexposed.

Figure C.3(c) shows the effect of subsequent chemical processing of the negative film. Note that a negative image, having colors complementary to those of the stimuli, has been formed. Where there was red exposure, there is cyan dye; where there was green exposure, there is magenta dye; and where there was blue exposure, there is yellow dye. Figure C.3(d) shows the resulting colors on the negative film.

Figure C.3(e) shows the result of exposing the negative film onto an appropriate photographic print medium and subsequently processing that print medium. Photographic print media intended for this purpose are also *negative working*, i.e., the amount of image dye they form increases with increased exposure of the medium. The result of exposing one negative medium onto another negative medium is a positive image, as shown in Figure C.3(f). Print media having a paper support are used to make reflection-print images, and media having a clear support are used to make images for projection or back illumination.

Image-dye formation in photographic negative media results from two basic chemical reactions. First, a negative, black and white, silver image is formed when exposed silver halide grains are chemically developed. The unexposed grains do not develop. In this development reaction, exposed silver halide (Ag X_{exp}) is reduced by the developing agent to form metallic silver (Ag °), a halide ion (X^-) is produced, and the developing agent itself is oxidized:

$$\begin{aligned} &AgX_{exp} + \text{Developer} \rightarrow \\ &Ag^o + X^- + \text{Oxidized developer} \end{aligned} \tag{C.1}$$

Special developing agents are used such that the resulting oxidized developer is "half" of a dye molecule. The other "half" of the dye molecule is the coupler mentioned earlier. Different couplers are incorporated in the red-light-sensitive, green-light-sensitive, and blue-light-sensitive layers such that cyan, magenta, and yellow image dyes are produced, respectively. These image dyes are formed by the reactions of oxidized developer with the

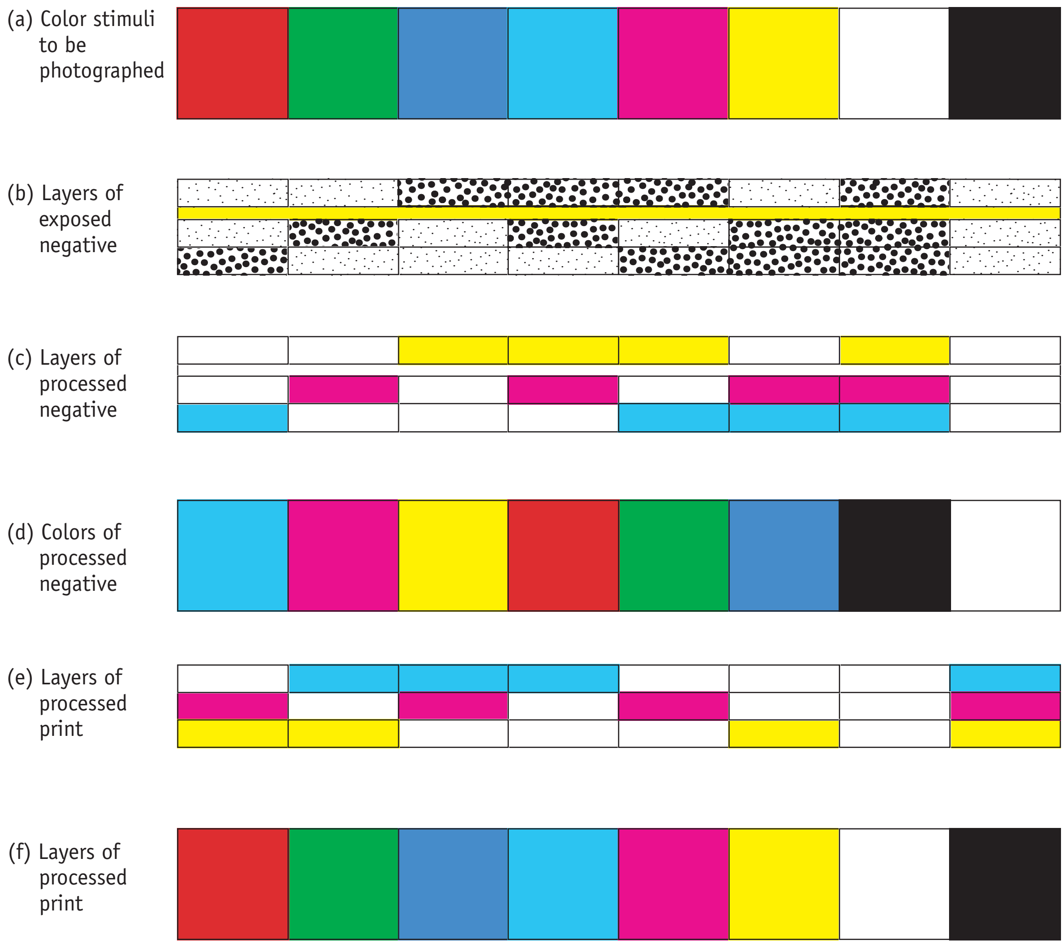


Figure C.3 Effect of various color stimuli on a photographic negative film, and the results of printing that film onto a photographic print medium.

respective couplers:

$$
\begin{aligned}
&\text{Oxidized developer}\\
&\quad + \text{Cyan-dye-forming coupler} \rightarrow \text{Cyan dye}\\
&\text{Oxidized developer}\\
&\quad + \text{Magenta-dye-forming coupler}\\
&\qquad \rightarrow \text{Magenta dye}\\
&\text{Oxidized developer}\\
&\quad + \text{Yellow-dye-forming coupler} \rightarrow \text{Yellow dye}
\end{aligned} \tag{C.2}
$$

In addition to their primary role of forming image dyes within each layer of a film, reactions (C1) and (C2) provide several opportunities to create desirable interactions among the layers. As explained in Part II, interactions among the color channels are required in virtually all color-imaging systems in order to achieve an appropriate level of overall color saturation and to produce other desirable colorimetric results. In photographic negative films and other silver-halide-based photographic media, chemical interactions called *interlayer effects* are used to produce these necessary color interactions. Various chemical techniques are used such that silver and/or image-dye formation in one film layer influences the amount of image dye that is formed in one or both of the other two layers.

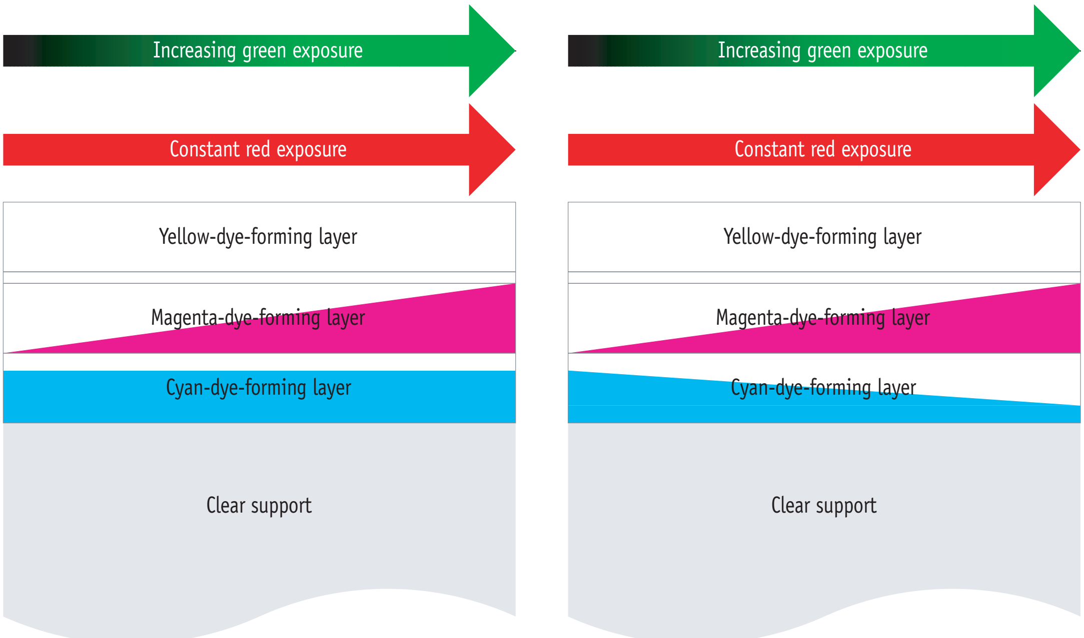


Figure C.4 Film cross-sections for two photographic negative films. The film on the left does not exhibit a chemical interlayer effect; the amount of cyan dye that forms in the red-light-sensitive layer is independent of the amount of magenta dye that forms above it. The film on the right does exhibit a chemical interlayer effect; the amount of cyan dye that forms is reduced in proportion to the amount of magenta dye that forms above it.

Figure C.4 illustrates the basic nature of an interlayer effect. The figure shows cross-sections of two photographic negative films. Both films have been exposed to green light that increases logarithmically in intensity from left to right. As a result of this exposure, the amount of magenta dye that forms in the green-light-sensitive layer of the processed film also will increase from left to right. In addition, both films have been exposed to red light of uniform intensity from left to right. The film on the left has no chemical interlayer interactions, so the amount of cyan dye that forms in the red-light-sensitive layer is uniform from left to right (like the light that exposed it). The film on the right shows a chemical interlayer interaction. Chemical by-products generated during the silver formation reaction (C1) and/or by the dye formation reaction (C2) travel to the red-light-sensitive layer and inhibit the formation of silver and/or image dye in that layer. The amount of cyan dye that forms in the red-light-sensitive layer therefore is reduced in proportion to the amount of magenta dye that forms above it.

The chemical interlayer interaction shown in the figure can be created by a number of different mechanisms. For example, the red-light-sensitive layer can be made responsive to the amount of halide ion created by silver development reaction (C1) in the green-light-sensitive layer. Also, a special coupler can be used in the green-light-sensitive layer so that a development inhibitor is released during the dye formation reaction (C2).

Why photographic negative films are orange

The overall orange cast of virtually all color photographic negative films results from another type of color-correction mechanism that primarily is used to compensate for the crosstalk that occurs in the transfer of color information from the negative to a print material. This crosstalk was discussed in some detail in Chapter 8. The compensating mechanism derives from the image-dye formation reaction (C2) discussed previously.

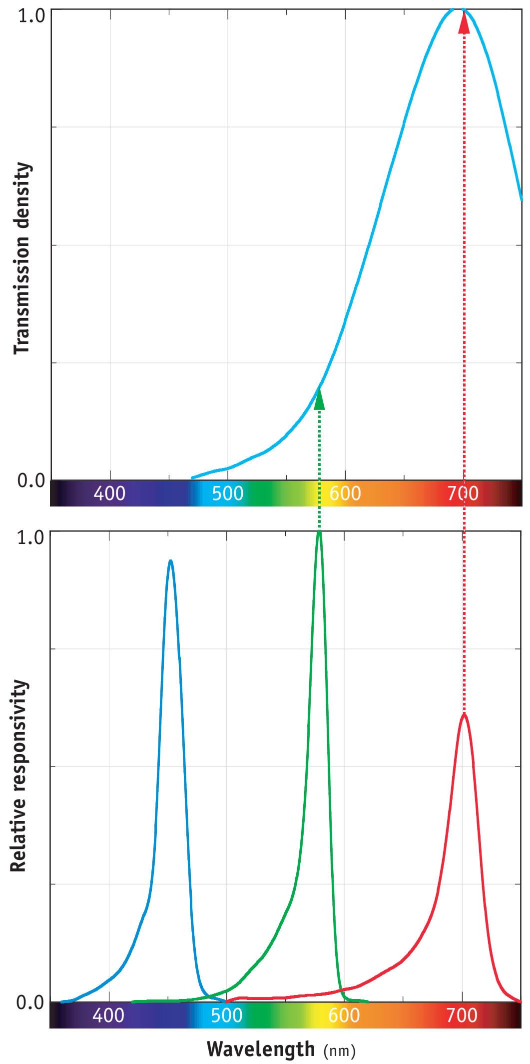


Figure C.5 Spectral characteristics of a cyan dye having unwanted green printing density, as measured according to the spectral responsivities shown below.

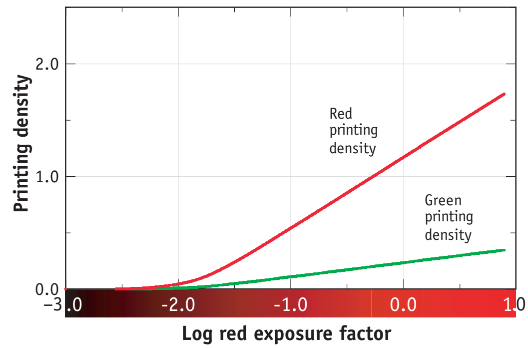


Figure C.6 As the amount of cyan dye in a negative increases, the red printing density will increase. The green printing density also will increase with the amount of cyan dye due to the unwanted absorption of green light by that cyan dye.

As that reaction progresses, coupler is used as image dye is formed. So if the coupler itself has some color to it, *that* color will go away as image dye is formed. This provides a means to compensate for certain unwanted spectral absorptions, as "seen" by the print material, of the image dye that is formed.

For example, suppose that in addition to having red printing density, the cyan dye that forms also has unwanted green printing density (Figure C.5). The more of this cyan dye that forms, the more unwanted green printing density there will be (Figure C.6). However, if the cyan-dye-forming coupler *itself* has green printing density (in other words, if the coupler is magenta colored, as shown in Figure C.7), *that* green printing density will go away as the coupler is used to form cyan image dye.

As Figure C.8 shows, where there is no red-light exposure, there will be no cyan dye, but there will be the green printing density of the unused magenta-colored coupler. Where there is *full* exposure, there will be a maximum amount of cyan dye, so there will be a maximum red printing density and a maximum amount of unwanted green printing density from that cyan dye. However, because it will have been used to form the cyan dye, no magenta-colored coupler will remain to contribute any green printing density. In areas where there is *partial* exposure, there will be green printing density from both the cyan dye and from any unused magenta-colored coupler.

If the appropriate amount of a coupler having the appropriate spectral absorption characteristics is used, the green printing density will stay *constant* as a function of red-light exposure (Figure C.9). The entire image therefore will have an overall magenta cast, but that cast can be compensated for by the use of more green light when the negative is printed. Because the colored coupler effectively will have corrected for the unwanted green absorption of the negative's cyan dye, a final print will look the same as a print made from a negative having a cyan dye with no unwanted green-light absorption.

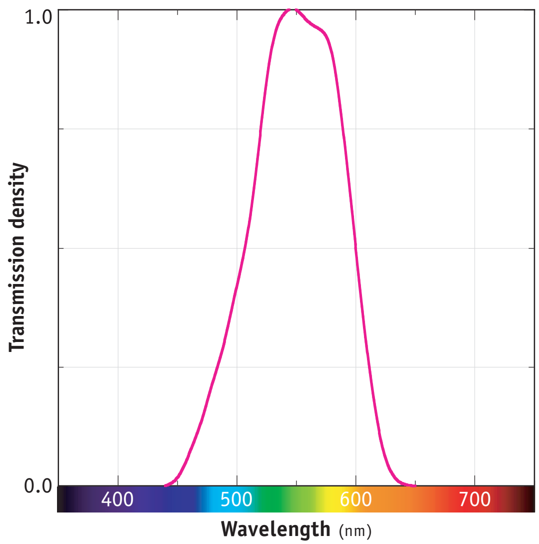


Figure C.7 Spectral transmission density of a cyan-dye-forming colored coupler. The coupler itself is magenta, so it absorbs green light. Its use compensates for the unwanted green-light absorption of the cyan dye.

In practice, colored couplers are incorporated in the cyan- and magenta-dye-forming layers of virtually all color-negative films. The colored coupler in the cyan layer is magenta-yellow; it corrects for the cyan image-forming dye's unwanted green and blue density. The colored coupler in the magenta layer is yellow; it corrects for the magenta image-forming

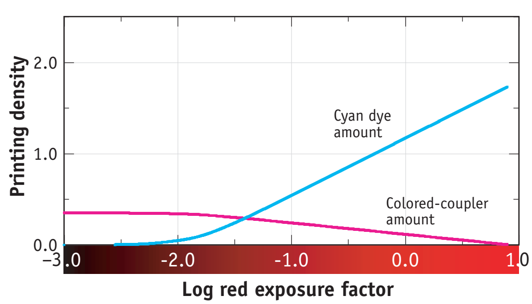


Figure C.8 At minimum red exposure, there is minimum cyan dye and maximum colored coupler. At maximum red exposure, there is maximum cyan dye and minimum colored coupler. At intermediary exposures, cyan dye and unused colored coupler both are present.

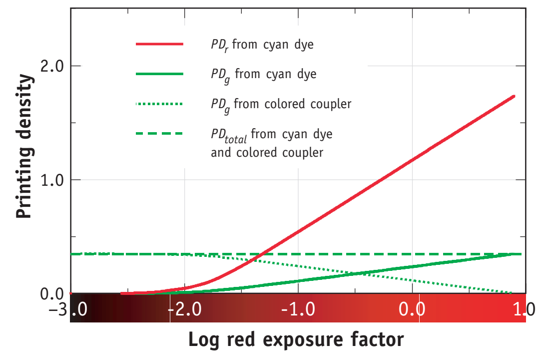


Figure C.9 When a magenta-colored cyan-dye-forming coupler is used in a color-negative film, the green printing density of that film becomes independent of the amount of cyan dye that is formed. The resulting overall magenta cast is accounted for in printing.

dye's unwanted blue density. It is the presence of these magenta-yellow- and yellow-colored couplers that gives negative films their characteristic overall orange cast.

Photographic transparency films

Figure C.10 shows a simplified cross-section of a photographic transparency film. Like a photographic negative film, a photographic transparency film has a blue-light-sensitive layer, a

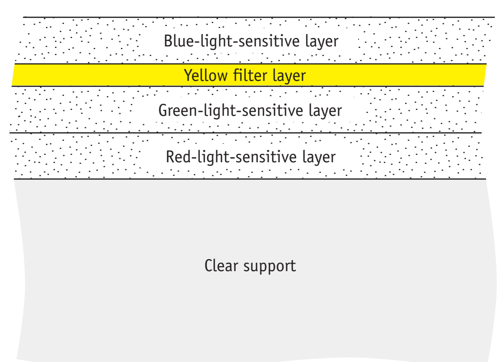


Figure C.10 Simplified cross-section of a color photographic transparency film (prior to chemical processing).

green-light-sensitive layer, and a red-light-sensitive layer coated on a clear support. Because the green-light-sensitive and red-light-sensitive layers also are inherently sensitive to blue light, a yellow filter layer must be coated above these layers to prevent blue light from reaching them. The red, green, and blue spectral sensitivities for a representative photographic transparency film are shown in Figure C.11.

When an exposed photographic transparency film is chemically processed, yellow, magenta, and cyan image dyes are formed in the blue-light-sensitive, green-light-sensitive, and red-light-sensitive layers, respectively. The nature of the chemical processing is such that a *positive* image results, i.e., the *maximum* amount of dye forms at the *minimum* exposure; the *minimum* amount of dye forms at the *maximum* exposure. Also during chemical processing, the yellow filter layer is made colorless.

Figure C.12(a) shows a series of color stimuli that are to be photographed by a photographic transparency film. Figure C.12(b) shows how each layer of the film is affected by exposure to the stimuli.

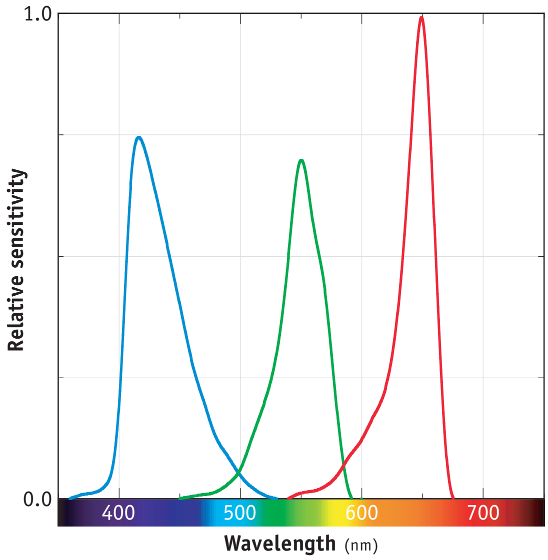


Figure C.11 Spectral sensitivities for a typical photographic transparency film.

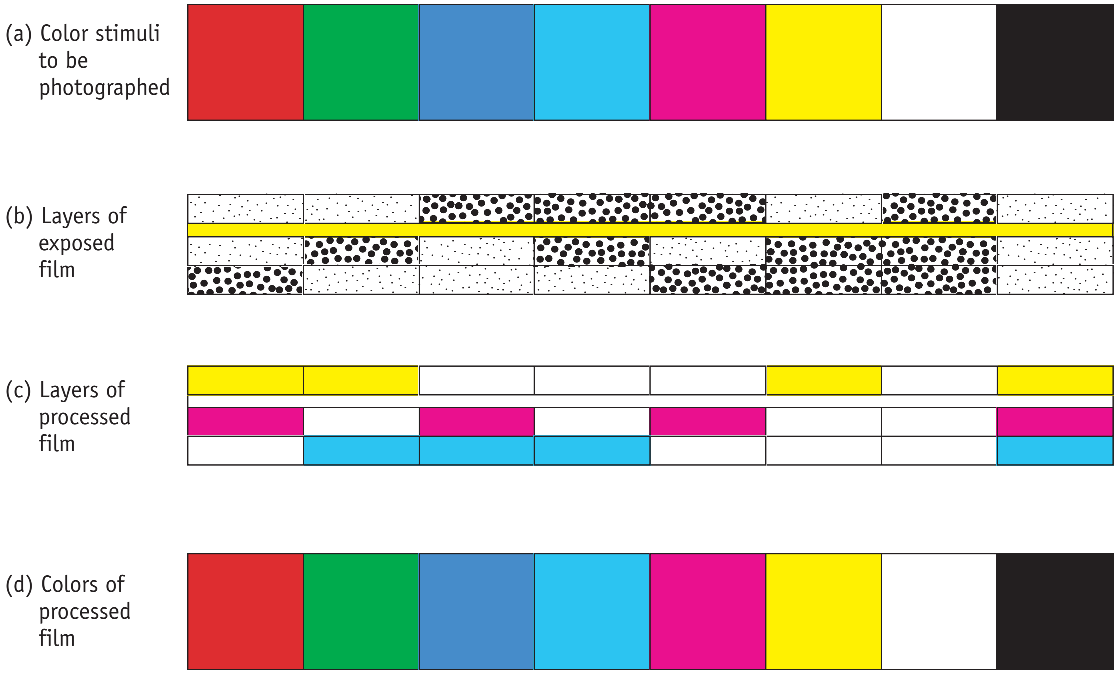


Figure C.12 Effect of various color stimuli on a photographic transparency film.

Figure C.12(c) shows the effect of subsequent chemical processing of the transparency film. In this processing, a negative silver image first is formed from the exposed silver halide. That image then is removed chemically, and a dye image is produced from the remaining (unexposed) silver halide. The result is a positive image having colors corresponding to those of the photographed stimuli, as shown in Figure C.12(d). For example, where there was red exposure, there is magenta and yellow dye (a red image); where there was green exposure, there is cyan and yellow dye (a green image); and where there was blue exposure, there is cyan and magenta dye (a blue image).

Because photographic transparency films usually are intended to be viewed directly, the couplers incorporated in these films must be colorless. Desirable interlayer interactions therefore must be created using chemical mechanisms, which generally are based on the silver development reaction (C1).

D

Adaptation

The term *adaptation* refers to various processes by which the visual mechanism adjusts to the conditions under which the eyes are exposed to radiant energy. The relationship between the physical characteristics of a given color stimulus and the perception of its color is strongly influenced by effects produced by a number of different forms of adaptation.

In this book, three types of visual adaptation of particular importance in color-imaging applications are considered: general-brightness adaptation, lateral-brightness adaptation, and chromatic adaptation.

General-Brightness Adaptation

General-brightness adaptation refers to the adjustments of the visual mechanism in response to the average luminance of the stimuli to which the eyes are exposed. General-brightness adaptation must be considered in the encoding of images displayed in viewing conditions in which the observer fully or partially adapts to the displayed image itself. Such adaptation will occur when images are viewed under conditions that eliminate or minimize other visual cues that might influence the adaptive state of the observer.

The following describes one procedure that can be used to transform the colorimetry of an image to *brightness-adapted* colorimetry. The procedure accounts for general-brightness adaptation effects.

First, the XYZ tristimulus values for each image pixel are determined, using standard colorimetric methods. Each Y tristimulus value then is adjusted by an experimentally determined scale factor, B, to form an adjusted tristimulus value, Y_b:

$$Y_b = BY \tag{D.1}$$

The value for the scale factor, B, will vary depending on the conditions of the viewing environment. For a reflection image, viewed in a normal environment, the factor is 1.0. For a transmission image, projected in a darkened environment, the factor will be greater than 1.0. This means that the projected image appears brighter than would be indicated by its unadjusted colorimetric values.

The X and Z tristimulus values of each pixel then are scaled by the ratio of the Y_b tristimulus value to the unadjusted tristimulus value, Y. This equal scaling of the tristimulus values maintains the chromaticity of the pixel:

$$\begin{aligned} X_b &= X\frac{Y_b}{Y} \\ Z_b &= Z\frac{Y_b}{Y} \end{aligned} \tag{D.2}$$

The effect of a general-brightness adaptation transformation from a dark-projection environment to a normal environment is shown in the measured and adjusted grayscales of Figure D.1. The adjusted grayscale produced by the transformation does *not*

Digital Color Management: Encoding Solutions, Second Edition Edward J. Giorgianni and Thomas E. Madden
 ISBN: 978-0-470-51244-9

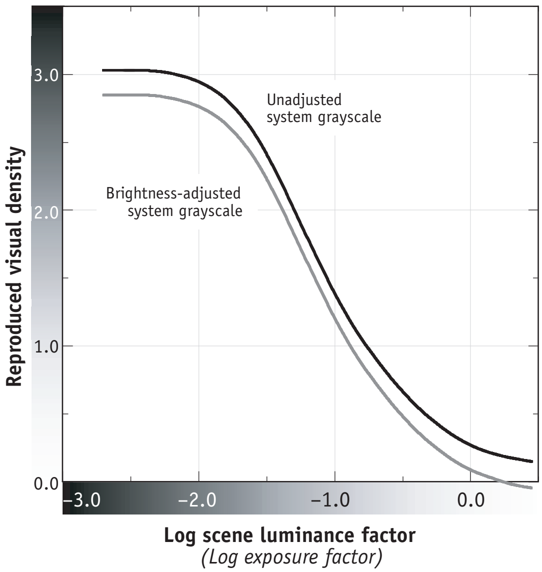


Figure D.1 The effect of a general-brightness adaptation transformation. The transformation determines the grayscale required to produce images in an average environment that match the relative brightnesses of images projected in a darkened environment.

describe what images on the measured film would look like if viewed in the normal environment. The adjusted grayscale instead defines the film grayscale that *would be required* in the normal environment to produce images that match the brightness of images produced when the measured film is projected in a darkened environment.

Note that the general-brightness transformation results in a downward shift in overall visual density; the photographic gamma (slope) of the curve is not affected by the transformation. However, the dark-surround viewing conditions leading to the shift in observer general-brightness adaptation also may induce *lateral*-brightness adaptation effects. These effects do require compensating photographic gamma adjustments, as discussed next.

Lateral-brightness adaptation

Lateral-brightness adaptation refers to changes in the sensitivities of areas of the retina as a function of light being received by adjacent receptors. One manifestation of lateral-brightness adaptation is that the perceived luminance contrast of an image is lowered if the areas immediately surrounding the image are relatively dark. This effect, sometimes called the *dark-surround effect*, occurs when an image, such as a photographic slide or motion picture, is projected in a darkened environment, and the viewing arrangement is such that the surround fills a substantial portion of the observer's field of view. A significant, but somewhat smaller, effect occurs in some displays of video images where the displayed image is viewed in a room that is dimly illuminated, but not completely dark. Again, the viewing arrangement must be such that the surround fills a substantial portion of the observer's field of view.

The following procedure is one method that can be used to transform the measured colorimetry of an image to *surround-adjusted* colorimetry. The procedure accounts for lateral-brightness adaptation effects.

First, the XYZ tristimulus values for each image pixel are determined, using standard colorimetric methods. Each Y tristimulus value then is modified by an experimentally determined power factor, S, to form a visually equivalent tristimulus value, Y_s, for the different surround:

$$Y_s = Y^S \tag{D.3}$$

The X and Z tristimulus values for the stimulus then are scaled by the ratio of the Y_s value to the unmodified tristimulus value, Y:

$$\begin{aligned} X_s &= X\frac{Y_s}{Y} \\ Z_s &= Z\frac{Y_s}{Y} \end{aligned} \tag{D.4}$$

This scaling maintains the chromaticity values (x, y) for the stimulus. In certain other lateral-brightness adaptation procedures, the chromaticity values are adjusted such that the chroma of the stimulus also is modified.

Applying a lateral-brightness adaptation transformation from darker-surround conditions to lighter-surround conditions ($S < 1.0$) results in a lowering in overall grayscale photographic gamma, as shown in Figure D.2. The meaning of this transformation is that an image having a grayscale of

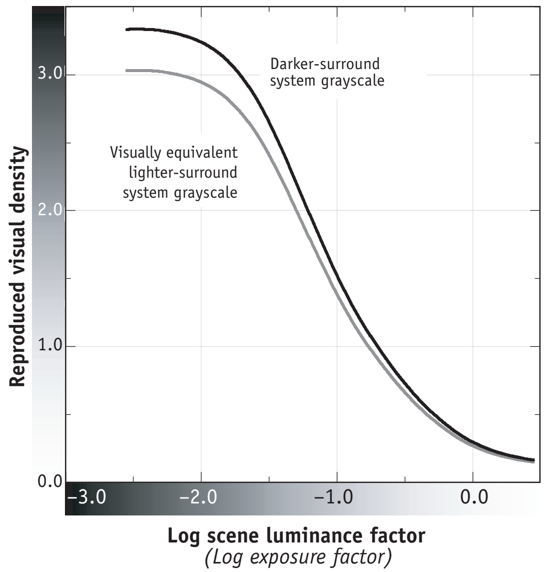


Figure D.2 Results of applying a lateral-brightness (or dark-surround) transformation. The transformation accounts for the fact that an image having a grayscale of lower slope would be required in lighter-surround conditions in order to match the luminance contrast of an image viewed in darker-surround conditions.

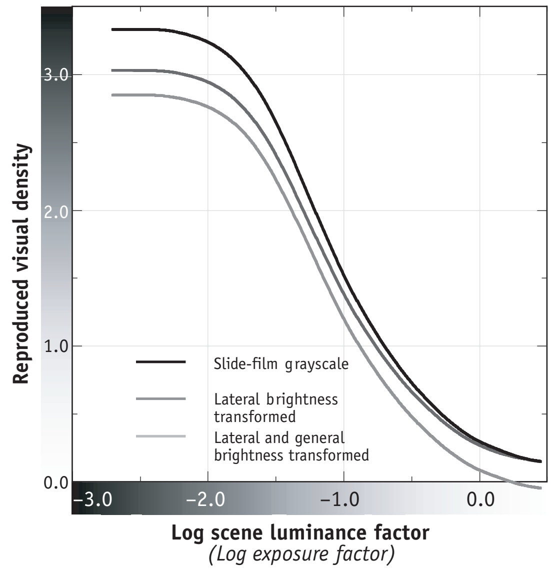


Figure D.3 Effects of applying lateral-brightness adaptation and general-brightness adaptation transformations to the grayscale visual density values of a photographic slide film.

lower photographic gamma would be required in lighter-surround conditions in order to match the luminance contrast of an image viewed in darker-surround conditions.

The value for the power factor, S, will vary depending on the particular conditions of the viewing environment. The magnitude of the dark-surround effect depends on the luminance of the surround relative to the average luminance of the viewed image and on the proportions of the visual fields filled by the image and by the surround.

It should be noted that because viewing conditions that induce lateral-brightness adaptation effects likely will also induce shifts in general-brightness adaptation, the magnitudes for the power-factor value, S, and the scale-factor value, B, will be related. In practice, both values are best determined from visual judging experiments. Such experiments should be performed using the actual viewing conditions involved. Figure D.3 shows the effect of applying both brightness-adaptation transformations.

Observer chromatic adaptation

Chromatic adaptation refers to the adjustments of the visual mechanism in response to the average chromaticity of the stimuli to which the eyes are exposed. Changes in the chromatic adaptation state of an observer must be accounted for in color encoding by the use of appropriate chromatic adaptation transformations. A chromatic adaptation transformation determines visually equivalent tristimulus values of a color stimulus that would produce, for a standard observer at one state of chromatic adaptation, a visual match to another color stimulus viewed by a standard observer at a different state of chromatic adaptation.

Although different types of chromatic adaptation transformations are available, their basic methodology generally is similar. First, XYZ tristimulus values, corresponding to a first state of adaptation, are transformed into three cone-response values, ρ, γ, and β. Next, the cone-response values are adjusted to account for the change in the observer's state of chromatic adaptation from the first state to a second state. Finally, the adjusted cone-response values are transformed to produce visually equivalent *XYZ* tristimulus values for the second state of chromatic adaptation.

In this book, von Kries transformations, in which cone-response values are adjusted by linear scaling, are used for all chromatic adaptation transformations. The following discussion and example illustrate the basic procedure.

In this example, a von Kries transformation from adaptive white chromaticity coordinate values $x = 0.3127$, $y = 0.3290$ (D_{65}) to the adaptive white chromaticity coordinate values $x = 0.3457$, $y = 0.3585$ (D_{50}) will be derived.

First, the *XYZ* values for both adaptive white chromaticities must be determined. By convention, the Y value for a perfect white is set equal to 100. The values for X and Z can be calculated from the x, y chromaticity coordinates, using the following relationships:

$$\begin{aligned} x + y + z &= 1 \\ \therefore z &= 1 - x - y \end{aligned} \tag{D.5a}$$

$$\begin{aligned} X &= Y\frac{x}{y} = 100\frac{x}{y} \\ Z &= Y\frac{z}{y} = 100\frac{z}{y} \end{aligned} \tag{D.5b}$$

For adaptive white chromaticity coordinates of $x = 0.3127$, $y = 0.3290$ (D_{65}):

$$z = 1 - 0.3127 - 0.3290 = 0.3583$$

$$X = 100\frac{0.3127}{0.3290} = 95.05$$

$$Y = 100$$

$$Z = 100\frac{0.3583}{0.3290} = 108.91$$

For adaptive white chromaticity coordinates of $x = 0.3457$, $y = 0.3585$ (D_{50}):

$$\begin{aligned} z &= 1 - 0.3457 - 0.3585 = 0.2958 \\ X &= 100\frac{0.3457}{0.3585} = 96.43 \\ Y &= 100 \\ Z &= 100\frac{0.2958}{0.3585} = 82.51 \end{aligned}$$

Next, the ρ, γ, and β cone-response values are determined from the *XYZ* values. The following von Kries matrix can be used for these computations:

$$\begin{bmatrix} \rho \\ \gamma \\ \beta \end{bmatrix} = \begin{bmatrix} 0.40024 & 0.70760 & -0.08081 \\ -0.22630 & 1.65320 & 0.04570 \\ 000.91822 & & \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} \tag{D.6}$$

For D_{65}:

$$\begin{aligned} \rho_{D_{65}} &= 0.40024(95.05) + 0.70760(100) \\ &\quad - 0.08081(108.91) = 100 \\ \gamma_{D_{65}} &= -0.22630(95.05) + 1.65320(100) \\ &\quad + 0.04570(108.91) = 100 \\ \beta_{D_{65}} &= 0.91822(108.91) = 100 \end{aligned}$$

For D_{50}:

$$\begin{aligned} \rho_{D_{50}} &= 0.40024(96.43) + 0.70760(100) \\ &\quad - 0.08081(82.51) = 102.69 \\ \gamma_{D_{50}} &= -0.22630(96.43) + 1.65320(100) \\ &\quad + 0.05470(82.51) = 98.48 \\ \beta_{D_{50}} &= 0.91822(82.51) = 75.76 \end{aligned}$$

Visually equivalent tristimulus values $XYZ_{D_{50}}$ for a color stimulus that would produce, for a standard observer chromatically adapted to the chromaticity of D_{50}, a visual match to another color stimulus, having tristimulus values $XYZ_{D_{65}}$ viewed by a standard observer who is chromatically adapted to the chromaticity of D_{65}, can be determined from the von Kries matrix (D.6), the ratios $(\rho_{D_{50}}/\rho_{D_{65}})$, $(\gamma_{D_{50}}/\gamma_{D_{65}})$, and $(\beta_{D_{50}}/\beta_{D_{65}})$ the inverse von Kries matrix as follows:

$$\begin{bmatrix} X_{D_{50}} \\ Y_{D_{50}} \\ Z_{D_{50}} \end{bmatrix} = [D6]^{-1} \begin{bmatrix} \rho_{D_{50}}/\rho_{D_{65}} & 0 & 0 \\ 0 & \gamma_{D_{50}}/\gamma_{D_{65}} & 0 \\ 0 & 0 & \beta_{D_{50}}/\beta_{D_{65}} \end{bmatrix} \begin{bmatrix} X_{D_{65}} \\ Y_{D_{65}} \\ Z_{D_{65}} \end{bmatrix} \tag{D.7a}$$

These mathematical operations can be combined to form a single chromatic adaptation transformation matrix:

$$\begin{bmatrix} X_{D_{50}} \\ Y_{D_{50}} \\ Z_{D_{50}} \end{bmatrix} = \begin{bmatrix} 1.0161 & 0.0553 & -0.0522 \\ 0.0060 & 0.9956 & -0.0012 \\ 0.0000 & 0.0000 & 0.7576 \end{bmatrix} \begin{bmatrix} X_{D_{65}} \\ Y_{D_{65}} \\ Z_{D_{65}} \end{bmatrix} \tag{D.7b}$$

E

Viewing Flare

When an image is viewed, the observer ideally should see only the light from the image itself. But in most actual viewing conditions, the observer also will see *flare light*, i.e., stray light containing no image information. For example, when viewing a video image displayed on a monitor, the observer may see flare light reflected from the glass faceplate of that monitor. The light causing the flare might come from ordinary light sources in the viewing environment (overhead lamps, window light, etc.). It also might come from the video display itself—if, for example, a workstation user is wearing a white shirt that reflects monitor light back to the CRT faceplate.

In reflection-image viewing, flare occurs as light is reflected directly from the front surface of the medium, without passing through the colorants that make up the image. The amount of flare will vary, depending on the surface texture of the medium, viewing angle, and other factors. In projection-image viewing, viewing flare can be caused by stray projector light. Flare light also can come from projected light that first reflects from the projection screen to the various surfaces in the room and then reflects back to the screen. The amount of flare will depend on the type of screen, the characteristics of the room, the viewing angle, and other factors.

Effect of viewing flare on image grayscale

Viewing flare adds light, essentially uniformly, to the entire image being viewed. Flare light will significantly brighten darker areas of an image, but it will have a less apparent effect on brighter areas of the image. As a result, the luminance contrast of the image will be reduced.

The effect of flare light on the grayscale of an imaging medium can be computed from a measurement of the grayscale in the absence of viewing flare and a measurement of the relative amount of flare light (Figure E.1). For example, if the flare light in a reflection-print viewing environment is equal to 1 %, referenced to the amount of light reflected from the print's reproduction of a perfect white, the computations can be performed as follows:

1. The density corresponding to the reproduction of a perfect white, D_w, is determined. For the example of Figure E.1, D_w is 0.12.
2. D_w is converted to a linear reflectance value:

$$R_w = 10^{-D_w} \tag{E.1}$$

In this example, where D_w is 0.12,

$$R_w = 10^{-0.12} = 0.7586$$

Digital Color Management: Encoding Solutions, Second Edition Edward J. Giorgianni and Thomas E. Madden
 ISBN: 978-0-470-51244-9

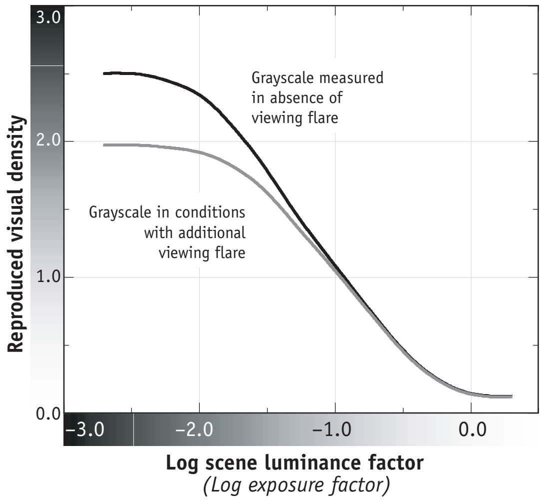


Figure E.1 Effect of viewing flare on a system grayscale characteristic.

3. The amount of flare light, fl, relative to the amount of light from the reproduced white, is determined from the percent flare, pf, and the white reflectance, R_w:

$$fl = \frac{pf}{100} R_w \tag{E.2}$$

In this example, where the percent flare is 1.0 %,

$$fl = \left(\frac{1.0}{100}\right) 0.7586 = 0.007\,586$$

4. For each point on the grayscale characteristic, which is measured in the absence of flare light, flareless density values are converted to flareless reflectance values:

$$R_{nf} = 10^{-D_{nf}} \tag{E.3}$$

For instance, at a point where the density (flareless), D_{nf}, is 2.0,

$$R_{nf} = 10^{-2.0} = 0.01$$

5. Flare then is added to each flareless reflectance value of the curve:

$$R_f = R_{nf} + fl \tag{E.4}$$

In this example, at a point where the density (flareless) is 2.0,

$$R_f = 0.01 + 0.007\,586 = 0.017\,586$$

6. The computed reflectance (with flare) values are converted to density (with flare) values:

$$D_f = -\log_{10}(R_f) \tag{E.5}$$

In this example, at a point where the density (no flare) is 2.0,

$$D_f = -\log_{10}(0.017\,586) = 1.7548$$

At a point where the density (no flare) is 1.0,

$$D_f = 0.9682$$

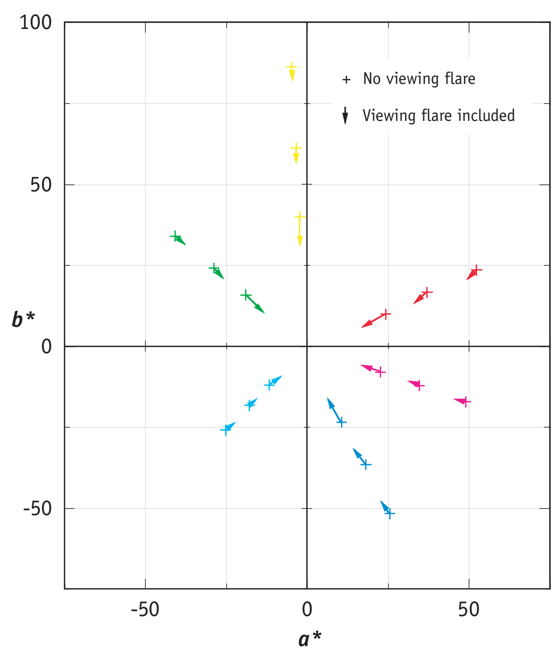


Figure E.2 Effect of viewing flare on colors. The addition of flare light moves colors toward the color of the flare light itself. Colors will be lower in chroma, as can be seen here. They also will be lighter and therefore less saturated.

The 1 % flare grayscale shown in Figure E.1 was computed by applying this procedure to all values of the flareless grayscale.

Colorimetric effects

The colorimetric effects of viewing flare can be computed using the principles of additive color mixing. The tristimulus values for a stimulus (with flare) can be determined by adding appropriate amounts of the tristimulus values for the flare light itself to the tristimulus values for the stimulus (without flare). As shown in Figure E.2, the addition of flare light moves colors toward the color of flare light itself.

The first step in the computation is to determine the tristimulus values, X_{fl}, Y_{fl}, and Z_{fl}, for the flare light, based on the specification of that light in terms of its chromaticity coordinates x_{fl}, y_{fl}, and z_{fl}, and its percentage, pf, of the image white luminance, Y_w:

$$
\begin{aligned}
Y_{fl} &= \frac{pf}{100} Y_w \\
X_{fl} &= \frac{x_{fl}}{y_{fl}} Y_{fl} \\
Z_{fl} &= \frac{z_{fl}}{y_{fl}} Y_{fl}
\end{aligned}
\tag{E.6}
$$

For each color stimulus, a modified stimulus then can be computed by adding the tristimulus values for the flare light to the tristimulus values for the color stimulus, as shown below:

$$
\begin{aligned}
X_f &= X_{nf} + X_{fl} \\
Y_f &= Y_{nf} + Y_{fl} \\
Z_f &= Z_{nf} + Z_{fl}
\end{aligned}
\tag{E.7}
$$

where X_{nf}, Y_{nf}, and Z_{nf} are the tristimulus values for a stimulus with no flare light, and X_f, Y_f, and Z_f are the tristimulus values for a stimulus with flare light added. The effect of this stimulus modification is shown in Figure E.2.

F

Scene-Based Color Encoding Specifications

The fundamental characteristics of scene-based color encoding specifications often make them the best choice for representing color within an imaging system. In particular, scene-based color encoding specifications are well suited for representing color at the initial stages of imaging systems where editing, compositing, and other forms of image manipulation are to be performed. Two representative scene-based color encoding specifications, Kodak PhotoYCC color interchange space and RIMM/ERIMM, are described in this appendix.

Kodak PhotoYCC color interchange space

Although designed and developed specifically for the Kodak Photo CD system, Kodak PhotoYCC color interchange space became widely used in numerous professional applications for the interchange of color images. This is understandable because its principal features—large color gamut and extensive luminance dynamic range, support for image compression, minimal quantization effects, fast output to softcopy displays—are important in professional imaging applications. Equally important is the fact that colors encoded in terms of PhotoYCC space values essentially represent original-scene colorimetry. However, because it is a fully defined appearance-based color encoding specification that includes specified encoding reference viewing conditions, Kodak PhotoYCC color interchange space also can be used to represent reproduced colors as well as other types of colors, regardless of their origin.

For color stimuli that are meant to be viewed in actual conditions that correspond to the encoding reference conditions, PhotoYCC space values are computed by applying a series of simple mathematical operations to standard CIE colorimetric values. For color stimuli that are meant to be viewed in actual conditions that differ from those of the encoding reference, it is necessary to include appropriate colorimetric transformations to determine visually equivalent CIE colorimetric values for the encoding reference conditions.

The colorimetric transformations must account for differences in the amount of viewing flare in the actual and encoding reference viewing conditions, as well as for alterations in observer perception that would be induced by other differences in those conditions. The visually equivalent standard CIE colorimetric values resulting from these transformations then are encoded in terms of PhotoYCC color interchange space values.

The specified conditions of the PhotoYCC space encoding reference viewing environment, listed in

Digital Color Management: Encoding Solutions, Second Edition Edward J. Giorgianni and Thomas E. Madden
 ISBN: 978-0-470-51244-9

Table F.1 PhotoYCC color encoding specification and color-encoding reference viewing conditions.

	Photo YCC Color Space
Colorimetry	1931 CIE flareless measurement
Data metric	PhotoYCC space $Y = (255/1.402)\ Luma$ $C_1 = 111.40\ Chroma_1 + 156$ $C_2 = 135.64\ Chroma_2 + 137$
Viewing conditions:	
Viewing flare	None
Image surround	Average
Luminance level	>1 600 cd/m^2
Adaptive white	$x = 0.3127$, $y = 0.3290$

Table F.1, correspond to those of typical outdoor scenes. These conditions are as follows:

- *Viewing flare* is specified as *none*; any flare light in an original-scene environment is part of the scene itself.
- The *surround type* is defined as *average*; scene objects typically are surrounded by other, similarly illuminated, scene objects.
- The *luminance level* is representative of typical daylight-illuminated objects. Note that this luminance level is at least an order of magnitude higher than that typical of indoor conditions.
- The chromaticity coordinates of the observer *adaptive white* are those of CIE Standard Illuminant D_{65}.

An adaptive white is defined here, and throughout this book, as a color stimulus that an adapted observer would judge to be perfectly achromatic and to have a luminance factor of unity. Although the chromaticity of the adaptive white most often will be that of the scene illuminant, it may be quite different in certain cases. For example, it may differ if the observer is only partially adapted to the illuminant. The adaptive white therefore defines only the chromatic adaptive state of the *observer*; it does *not* define the chromaticity, or the spectral power distribution, of the scene illuminant. The following examples provide additional details on this point and on PhotoYCC space color encoding.

Example 1: This example describes the PhotoYCC space encoding of a daylight-illuminated scene, recorded using the Photo CD reference image-capture device, when the adaptive white corresponds to D_{65} chromaticity.

The red, green, and blue spectral responsivities of the reference image-capture device, shown in Figure F.1, are equivalent to the color-matching functions for the reference primaries defined in Recommendation ITU-R BT.709. The CIE chromaticities for these red, green, and blue reference primaries, and for CIE Standard Illuminant D_{65}, are given in Table F.2.

Photo CD reference image-capture device RGB_{709} tristimulus values for the stimuli of a scene can be calculated using the spectral responsivities of the reference image-capture device:

$$\begin{aligned} R_{709} &= k_r \sum_{\lambda=380}^{780} S(\lambda)\, R(\lambda)\, \overline{r}(\lambda) \\ G_{709} &= k_g \sum_{\lambda=380}^{780} S(\lambda)\, R(\lambda)\, \overline{g}(\lambda) \\ B_{709} &= k_b \sum_{\lambda=380}^{780} S(\lambda)\, R(\lambda)\, \overline{b}(\lambda) \end{aligned} \tag{F.1}$$

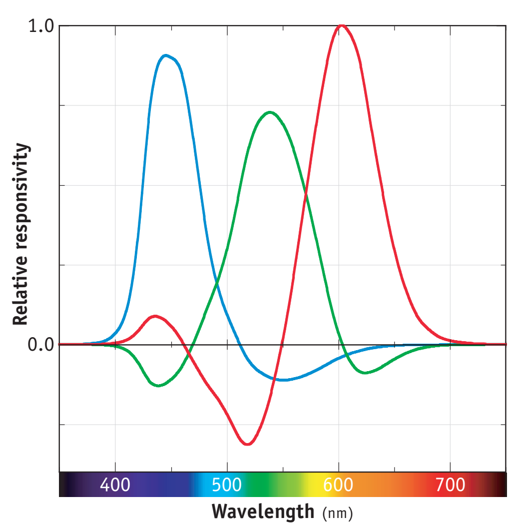


Figure F.1 Responsivities of the Photo CD reference image-capture device.

Table F.2 Chromaticities of ITU–R BT.709 Reference Primaries and CIE Standard Illuminant D_{65}.

	Red	Green	Blue	D_{65}
x	0.6400	0.3000	0.1500	0.3127
y	0.3300	0.6000	0.0600	0.3290
z	0.0300	0.1000	0.7900	0.3583
u'	0.4507	0.1250	0.1754	0.1978
v'	0.5229	0.5625	0.1579	0.4683

where R_{709}, G_{709}, and B_{709} are the reference image-capture device red, green, and blue tristimulus (exposure) values; $S(\lambda)$ is the spectral power distribution of a light source; $R(\lambda)$ is the spectral reflectance or transmittance of an object; $\overline{r}(\lambda)$, $\overline{g}(\lambda)$, and $\overline{b}(\lambda)$are the red, green, and blue spectral responsivities of the reference image-capture device; and k_r, k_g, *and* k_b are normalizing factors determined such that R_{709}, G_{709}, and $B_{709} = 1.00$ when the stimulus is a perfect white diffuser. By definition, all measurements made by the reference image-capture device are flareless, so it is not necessary to adjust the calculated RGB_{709} tristimulus values for instrument flare.

Since the red, green, and blue spectral responsivities of the reference image-capture device are linear combinations of the $\overline{x}(\lambda)$, $\overline{y}(\lambda)$, and $\overline{z}(\lambda)$ color-matching functions for the CIE 1931 Standard Colorimetric Observer, its RGB_{709} tristimulus values also can be calculated directly from CIE XYZ tristimulus values, using the following relationship:

$$\begin{bmatrix} R_{709} \\ G_{709} \\ B_{709} \end{bmatrix} = \begin{bmatrix} 3.2410 & -1.5374 & -0.4986 \\ 0.9692 & 1.8760 & 0.0416 \\ 0.0556 & -0.2040 & 1.0570 \end{bmatrix} \begin{bmatrix} \dfrac{X_{scene}}{100} \\ \dfrac{Y_{scene}}{100} \\ \dfrac{Z_{scene}}{100} \end{bmatrix} \tag{F.2}$$

where

$$\begin{aligned} X_{scene} &= k \sum_{\lambda=380}^{780} S(\lambda)\, R(\lambda)\, \overline{x}(\lambda) \\ Y_{scene} &= k \sum_{\lambda=380}^{780} S(\lambda)\, R(\lambda)\, \overline{y}(\lambda) \\ Z_{scene} &= k \sum_{\lambda=380}^{780} S(\lambda)\, R(\lambda)\, \overline{z}(\lambda) \end{aligned} \tag{F.3}$$

The value for the normalizing factor, k, in Equations (F.3) is determined such that $Y_{scene} = 100$ when the stimulus is a perfect white diffuser. In the PhotoYCC space encoding process, RGB_{709} tristimulus values less than zero and RGB_{709} tristimulus values greater than 1.00 are retained.

Reference image-capture device RGB_{709} tristimulus values next are transformed to nonlinear values, $R'G'B'_{709}$, as follows:

For R_{709}, G_{709}, $B_{709} \geq 0.018$:

$$\begin{aligned} R'_{709} &= 1.099R_{709}^{0.45} - 0.099 \\ G'_{709} &= 1.099R_{709}^{0.45} - 0.099 \\ B'_{709} &= 1.099R_{709}^{0.45} - 0.099 \end{aligned} \tag{F.4a}$$

For R_{709}, G_{709}, $B_{709} \leq -0.018$:

$$\begin{aligned} R'_{709} &= -1.099\,|R_{709}|^{0.45} + 0.099 \\ G'_{709} &= -1.099\,|G_{709}|^{0.45} + 0.099 \\ B'_{709} &= -1.099\,|B_{709}|^{0.45} + 0.099 \end{aligned} \tag{F.4b}$$

For $-0.018 < R_{709}$, G_{709}, $B_{709} < 0.018$:

$$\begin{aligned} R'_{709} &= 4.5R_{709} \\ G'_{709} &= 4.5G_{709} \\ B'_{709} &= 4.5B_{709} \end{aligned} \tag{F.4c}$$

The $R'G'B'_{709}$values then are transformed to luma and chroma values as follows:

$$\begin{bmatrix} Luma \\ Chroma_1 \\ Chroma_2 \end{bmatrix} = \begin{bmatrix} 0.299 & 0.587 & 0.114 \\ -0.299 & -0.587 & 0.886 \\ 0.701 & -0.587 & -0.114 \end{bmatrix} \begin{bmatrix} R'_{709} \\ G'_{709} \\ B'_{709} \end{bmatrix} \tag{F.5}$$

Finally, luma/chroma values are converted to digital code values Y, C_1, and C_2. For 24-bit (8 bits per channel) encoding, PhotoYCC space values are the nearest integers to the values determined from the following equations:

$$\begin{aligned} Y &= \frac{255}{1.402} Luma \\ C_1 &= 111.40\, Chroma_1 + 156 \\ C_2 &= 135.64\, Chroma_2 + 137 \end{aligned} \tag{F.6}$$

Example 2: This example describes the encoding of a daylight-illuminated scene captured using a reference image-capture device, or any other device having CIE color-matching-function spectral responsivities, when the observer adaptive white in the input-image viewing environment corresponds to a chromaticity *other than* that of D_{65}.

CIE *XYZ* tristimulus values for the illuminated objects of the scene are first computed, using standard colorimetric procedures. Because the observer adaptive white is other than that specified for the reference viewing environment, an appropriate chromatic adaptation transformation (such as a von Kries transformation, as described in Appendix D) must be applied to form visually equivalent tristimulus values for the PhotoYCC space reference viewing environment (D_{65} adaptive white chromaticity).

A chromatic adaptation transformation matrix can be determined from the chromaticity of the actual adaptive white and the specified chromaticity of the PhotoYCC space encoding reference adaptive white.

For example, if the actual adaptive white chromaticity coordinates correspond to those of D_{50}, the transformation matrix would be

$$\begin{bmatrix} X_{scene(D_{65})} \\ Y_{scene(D_{65})} \\ Z_{scene(D_{65})} \end{bmatrix} = \begin{bmatrix} 0.9845 & -0.0546 & 0.0678 \\ -0.0060 & 1.0048 & 0.0012 \\ 0.0000 & 0.0000 & 1.3200 \end{bmatrix} \begin{bmatrix} X_{scene(D_{50})} \\ Y_{scene(D_{50})} \\ Z_{scene(D_{50})} \end{bmatrix} \tag{F.7}$$

The $XYZ_{scene(D_{50})}$ tristimulus values next are transformed to reference image-capture device RGB_{709} tristimulus values, using Equation (F.2), and then are transformed to PhotoYCC space digital code values using Equations (F.4a) to (F.6).

Example 3: This example again describes the encoding of a daylight-illuminated original scene, but using an actual image-capture device (such as a digital still camera) having red, green, and blue spectral responsivities that are not equivalent to any set of CIE color-matching functions.

Original-scene RGB_{device} tristimulus values, which have been measured by the actual image-capture device, first must be corrected to compensate for any optical deficiencies, such as lens flare and spatial nonuniformities, of the device itself. The corrected RGB_{device} tristimulus values then are transformed to reference image-capture device RGB_{709} tristimulus values. This can be accomplished by a matrix transformation:

$$\begin{bmatrix} R_{709} \\ G_{709} \\ B_{709} \end{bmatrix} = [\boldsymbol{M}] \begin{bmatrix} R_{device} \\ G_{device} \\ B_{device} \end{bmatrix} \tag{F.8}$$

However, because the device has spectral responsivities that are not equivalent to a set of color-matching functions, this transformation cannot perfectly translate all possible original-scene stimuli. Therefore, appropriate statistical procedures must be used in the derivation of the transformation matrix. For example, the matrix may be derived such that the resulting colorimetric errors from a captured set of particularly important test colors are minimized. The remaining transformations, from the approximated RGB_{709} tristimulus values to PhotoYCC space values, are the same as in Example 1, Equations (F.4a) to (F.6).

Example 4: It sometimes is desirable to encode colors from reproduced images (such as reflection prints, transparencies, artwork, CRT displays, etc.) in terms of PhotoYCC space values. One procedure for doing this is to first encode the reproduced colors in terms of the color encoding specification (CES), or the identical color interchange specification (CIS), of the Unified Paradigm system described in Part IV. Numerous examples of this encoding are given in Chapters 20 and 21. The CES/CIS values then can be transformed to PhotoYCC space values using the procedures described in Appendix G.

RIMM RGB and ERIMM RGB color encoding specifications

RIMM RGB and ERIMM RGB are color encoding specifications designed for scene-based color encoding. The two specifications are defined in terms of the same set of RGB reference primaries and an identical reference viewing environment. They differ only in their numerical encodings. In other words, the two specifications employ the same color-encoding *method*, but use different encoding *data metrics*. Specifically, the ERIMM RGB data metric is capable of encoding a larger dynamic range of scene information.

RIMM/ERIMM RGB Primaries

The color primaries of RIMM RGB and ERIMM RGB have been selected to meet the following criteria:

- To include a color gamut sufficient to encompass the chromaticities of important colors, including all real-world surface colors.
- To efficiently encode color information with minimum quantization artifacts.
- To be well suited for the application of nonlinear tonescale transformations, particularly those required to produce rendered images appropriate for display.

The first two criteria are somewhat conflicting. When all other factors are equal, increasing the encodable gamut will produce correspondingly larger quantization errors. The RIMM/ERIMM primaries have been selected such that their chromaticity boundaries are sufficiently large to encode most colors of interest without being extended so far that unacceptable quantization errors are produced.

Figure F.2 shows the chromaticities of the RIMM/ERIMM RGB primaries. As the figure illustrates, these primaries fully encompass the gamut of real-world surface colors without devoting a lot of space to nonrealizable colors outside the spectrum locus. Also shown are the sRGB primaries, which are identical to the RGB primaries specified in ITU-R BT.709, *The HDTV Standard for the Studio and for International Programme Exchange*. It can be seen that the sRGB/ITU-709 color gamut does not include significant portions of the real-world surface-color gamut. In particular, it excludes many important colors near the yellow-to-red boundary of the spectrum locus. The default Adobe wide-gamut RGB gamut, also shown in the figure, excludes some of these colors as well.

Although numerous sets of primaries in addition to those of RIMM/ERIMM RGB also might meet the specified requirements for gamut and quantization, it is the third criterion that distinguishes RIMM/ERIMM RGB. As discussed throughout this book, the process of transforming an image from an unrendered scene state to a rendered display state often includes the application of a nonlinear transform

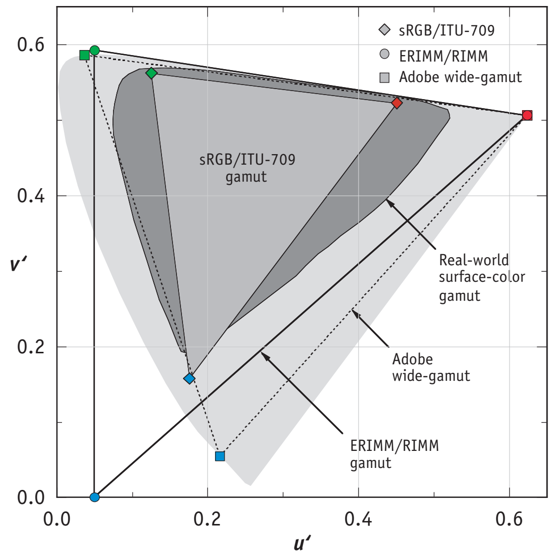


Figure F.2 Comparison of primaries in terms of *x*, *y* chromaticity coordinates.

to the individual color channels of the digital image. Figures F.3a and F.3b illustrate an example nonlinear transform, corresponding to a desired overall system grayscale, that might be used in transforming a scene-state image to a rendered-state image specifically for output on a reflection-print medium. Such transforms produce several desirable color and grayscale modifications, including:

- Increasing luminance contrast and color contrast in mid-tones and compressing the contrast of highlights and shadows.
- Increasing the chroma of in-gamut colors.
- Gamut mapping out-of-gamut colors in a simple but visually pleasing way.

However, nonlinear RGB transformations generally will also modify the *ratios* of the red, green, and blue image-color values. This can lead to undesirable hue shifts, especially in high-chroma colors, in the rendered-state image. Hue shifts are particularly undesirable when they occur in natural gradients within an image. Such gradients are common in the real world and occur, for example, when a

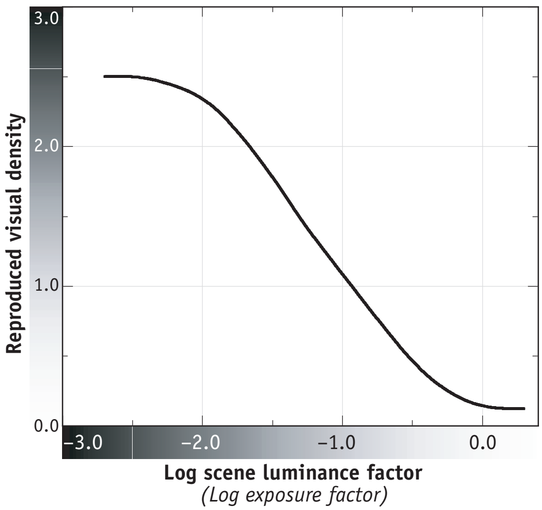


Figure F.3a A representative nonlinear transform, based on a desired overall system grayscale, for transforming scene-state image values to rendered-state image values for reflection-print output.

rounded object is illuminated by a moderately directional light source. In such situations, the chroma of the object color increases with distance from the highlight and then decreases again as the shadow deepens.

Although highlight-to-shadow hue shifts can be eliminated with appropriate three-dimensional signal processing, doing so complicates the transform development process and may require an unacceptable amount of additional image processing. Therefore an objective for the selection of the RIMM/ERIMM RGB primaries was to minimize objectionable hue shifts produced when one-dimensional nonlinear transformations, used by themselves, are applied to image data expressed in terms of the primaries.

In determining the RIMM/ERIMM primaries, hue shifts introduced by the application of nonlinear transformations were examined using computed highlight-to-shadow series for hundreds of colors. An example of such a series is shown in Figure F.4. Representative hue shifts associated with the selected RIMM/ERIMM primaries are shown in Figure F.5. The figure shows a series of line segments connecting the a^*, b^* values before and after a nonlinear transformation was applied to a highlight-to-shadow series in each of eight color directions. It can be seen that while small hue shifts are introduced for the most saturated colors in the blue and cyan

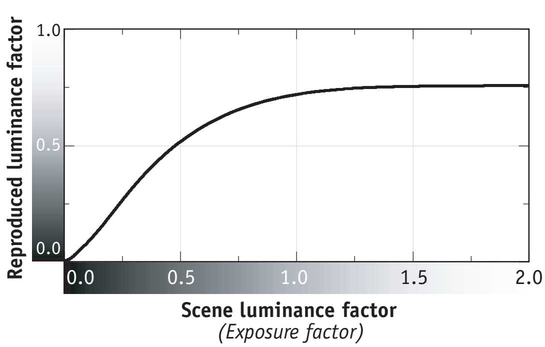


Figure F.3b The example transform of F.3a expressed in terms of linear values.

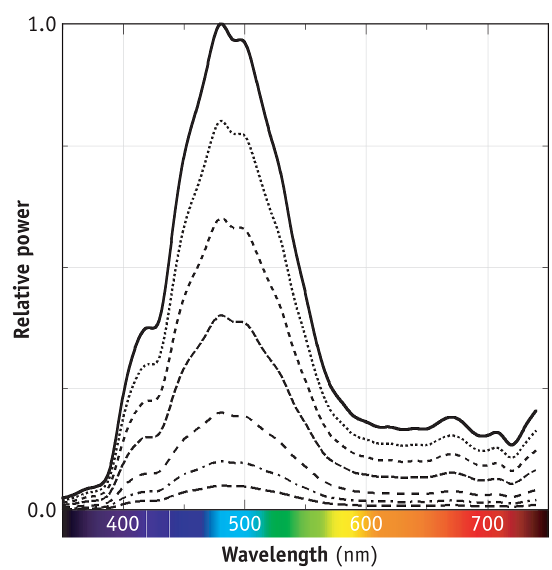


Figure F.4 Spectral power distributions for a representative shadow-to-highlight series of a single object.

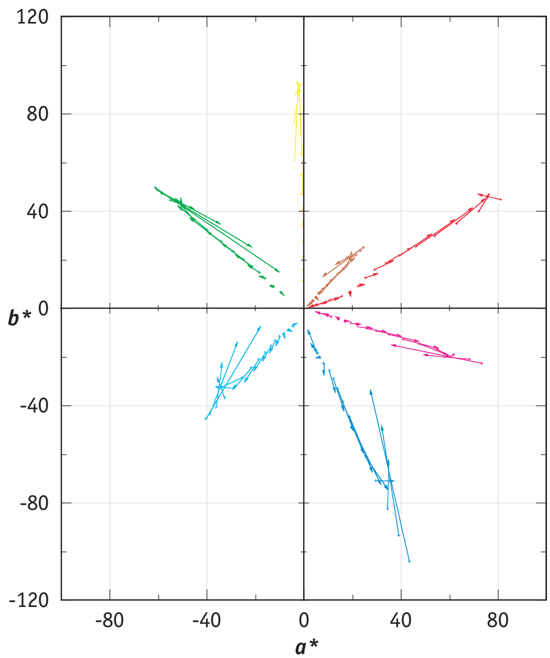


Figure F.5 Hue shifts for the RIMM/ERIMM RGB color encoding resulting from a representative nonlinear tonescale transform.

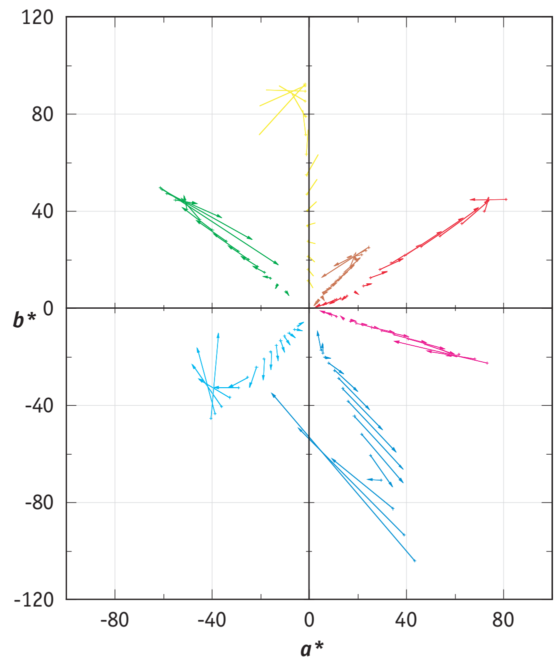


Figure F.6 Hue shifts for the default Adobe Photoshop wide-gamut RGB color-space color encoding resulting from a representative nonlinear tonescale transform.

directions, hue shifts elsewhere are quite small. The resulting hue shifts associated with the primaries of the default Adobe wide-gamut RGB color space are shown in Figure F.6. It can be seen that the hue shifts for these primaries generally are significantly larger than those of the RIMM/ERIMM primaries.

In the RIMM/ERIMM RGB encoding, scene color values are assumed to be encoded from flareless (or flare-corrected) colorimetric measurements based on the CIE 1931 Standard Colorimetric Observer. The relationship of RIMM/ERIMM RGB primary values to CIE XYZ values is as follows:

$$\begin{bmatrix} R_{RIMM} \\ G_{RIMM} \\ B_{RIMM} \end{bmatrix} = \begin{bmatrix} 1.3460 & -0.2556 & -0.0511 \\ -0.5446 & 1.5082 & 0.0205 \\ 0.0000 & 0.0000 & 1.2123 \end{bmatrix} \begin{bmatrix} \dfrac{X_{scene}}{100} \\ \dfrac{Y_{scene}}{100} \\ \dfrac{Z_{scene}}{100} \end{bmatrix} \tag{F.9}$$

To further define the color encoding specifications for RIMM RGB and ERIMM RGB, a set of encoding reference viewing conditions must be specified. Consistent with scene-based encoding, RIMM/ERIMM reference conditions are representative of typical outdoor environments and are defined as follows:

- Scene luminance level is greater than 1600 cd/m2.
- Viewing surround type is average, i.e., the overall luminance level and chromaticity of the surround are assumed to be similar to that of the scene.
- There is no viewing flare assumed for the scene, other than that which is already represented in the measured colorimetric values.
- The observer adaptive white point is specified by the chromaticity values for CIE Standard Illuminant D_{50} ($x = 0.3457$ and $y = 0.3585$).

Digital Encoding of RIMM RGB

Because the luminance dynamic range of scene-based encodings can be extensive, some form of nonlinear encoding must be employed at most practical bit depths in order to avoid visible quantization artifacts. A modified power-law encoding is used for RIMM RGB. A linear segment is included for the very lowest exposure values to overcome the noninvertibility of the very high slope of the power curve at such values. The encoding was defined such that the linear and nonlinear segments match in both value and derivative at the boundary. In equation form, this encoding is represented as follows:

$$C'_{RIMM} = \begin{cases} 0; & C_{RIMM} < 0.0 \\ \dfrac{I_{max}}{1.402} 4.5 C_{RIMM}; & 0.0 \le C_{RIMM} < 0.018 \\ \dfrac{I_{max}}{1.402} \left(1.099 C_{RIMM}^{0.45} - 0.099\right); & 0.018 \le C_{RIMM} < 2.0 \\ I_{max}; & C_{RIMM} \ge 2.0 \end{cases} \tag{F.10}$$

where C is either R, G, or B; I_{max} is the maximum integer value used for the nonlinear encoding; 2.0 is the maximum encoded exposure level; and

$$1.402 = 1.099 \left(2.0^{0.45}\right) - 0.099 \tag{F.11}$$

For the baseline 8-bit per channel RIMM RGB configuration, I_{max} is 255. In some applications, it may be desirable to use a higher bit precision version of RIMM RGB to further minimize quantization errors. For that reason, 12-bit per channel and 16-bit per channel versions of RIMM RGB also are defined. The only difference in the digital encodings is that the value of I_{max} is set to 4095 or 65535, respectively. When it is necessary to identify a specific precision level, the appropriate notation RIMM8 RGB, RIMM12 RGB, or RIMM16 RGB is used.

Inverse Encoding for RIMM RGB

To convert from *RIMM RGB* values back to the corresponding scene colorimetric values, it is only necessary to invert the nonlinear encoding

$$C_{RIMM} = \begin{cases} \dfrac{1.402}{I_{max}} \dfrac{C'_{RIMM}}{4.5}; & 0 \le C'_{RIMM} < \dfrac{0.081 I_{max}}{1.402} \\ \left(\dfrac{\frac{1.402}{I_{max}} C'_{RIMM} + 0.099}{1.099}\right)^{\frac{1}{0.45}}; & \dfrac{0.081 I_{max}}{1.402} \le C'_{RIMM} < I_{max} \end{cases} \tag{F.12}$$

and then apply the inverse matrix:

$$\begin{bmatrix} X_{scene} \\ Y_{scene} \\ Z_{scene} \end{bmatrix} = \begin{bmatrix} 79.76 & 13.52 & 3.13 \\ 28.80 & 71.19 & 0.01 \\ 0.00 & 0.00 & 82.49 \end{bmatrix} \begin{bmatrix} R_{RIMM} \\ G_{RIMM} \\ B_{RIMM} \end{bmatrix} \tag{F.13}$$

ERIMM RGB

The RIMM RGB color space is defined to have a luminance dynamic range that can encode information up to 200 % of the exposure value associated with a normally exposed perfect (100 %) diffuse white reflector in the scene. Although this range is adequate for some applications, it is insufficient to encode the full range of scene information captured by photographic negative films and by some types of digital cameras. For such cases, a variation of the RIMM RGB color space, referred to as Extended Reference Input Medium Metric RGB (ERIMM RGB), has been defined.

As with RIMM RGB, ERIMM RGB is representative of the colorimetry of an original scene. Accordingly, the two specifications share a common reference viewing environment. However, the nonlinear encoding function of ERIMM RGB is different from that of RIMM RGB. For ERIMM RGB, the design objectives were to increase the luminance dynamic range that can be represented and also to reduce the quantization interval size. In order to meet both objectives simultaneously, a minimum bit precision of 12 bits per color channel is used for ERIMM RGB.

Digital Encoding for ERIMM RGB

A modified logarithmic encoding is used for ERIMM RGB. A linear segment is included for the very lowest exposure values to overcome the non-invertibility of a logarithmic encoding at such values. The encoding was defined such that the linear and logarithmic segments match in both value and derivative at the boundary. In equation form, this encoding is represented as follows:

$$C'_{ERIMM} = \begin{cases} 0; & C_{RIMM} < 0.0 \\ I_{max}\left(\dfrac{78.9626}{e}\right) C_{RIMM}; & 0.0 \leq C_{RIMM} < e/1000 \\ I_{max}\left(\dfrac{\log_{10}(C_{RIMM}) + 3.0}{5.5}\right); & e/1000 \leq C_{RIMM} < 10^{2.5} \\ I_{max}; & C_{RIMM} \geq 10^{2.5} \end{cases} \tag{F.14}$$

where C is R, G, or B; I_{max} is the maximum integer value used for the nonlinear encoding; $10^{2.5}$ is the upper exposure limit that gets mapped to I_{max}; and $e/1000$ is the breakpoint between the linear and logarithmic segments, e being the base of the natural logarithm. For a 12-bit encoding, I_{max} is 4095, and for a 16-bit encoding I_{max} is 65535. In cases in which it is necessary to identify a specific precision level, the notation ERIMM12 RGB or ERIMM16 RGB is used.

The nonlinear function given in Equation (F.14) can be inverted to determine an inverse *ERIMM RGB* encoding function:

$$C_{ERIMM} = \begin{cases} \dfrac{e}{78.9626\,(2^n - 1)} C'_{ERIMM}; & 0 \leq C'_{ERIMM} < 0.078\,962\,6\,(2^n - 1) \\ antilog_{10}\left(\dfrac{5.5}{(2^n - 1)} C'_{ERIMM} - 3\right); & 0.078\,962\,6\,(2^n - 1) \leq C'_{ERIMM} < (2^n - 1) \end{cases} \tag{F.15}$$

where C'_{ERIMM} and C_{ERIMM} are the nonlinear and linear *ERIMM RGB* values, respectively, and as before C is R, G, or B.

G

Transformations for Color Interchange

Transformations for color interchange can be developed for any pair of systems having fully defined color encoding specifications. For *system-specific* interchange, various restricted interchange methods and data metrics can be used. For *general* color interchange, the communication of color information must be unrestricted, i.e., the interchange method itself must not impose limitations on luminance dynamic range, color gamut, or color interpretation.

In this appendix, several example transformations based on both restricted and unrestricted color interchange will be developed. Many of the examples will link Kodak PhotoYCC color interchange space to other color encoding specifications (Figure G.1). This particular interchange space was selected for inclusion here because it is representative of scene-based encoding and because its properties support both general and system-specific color interchange. Other important transformations using the color encoding specification developed for the example color-management systems described in Chapters 20 and 21 also will be described. Figure G.2 illustrates the colorimetric relationships of original scenes, the various color encoding specifications, and reproduced images. The characteristics of the individual color encoding specifications are as follows:

PhotoYCC space: Kodak PhotoYCC color interchange space was used for color encoding on the Kodak Photo CD system. A number of transforms to and from the space are given here for two principal reasons. First, there currently are billions of PhotoYCC files in image archives throughout the world, and there is considerable interest in transforming these files to other formats. Second, PhotoYCC space is a scene-based color encoding, and the example transforms illustrate the general principles involved in transforming between scene-based and rendered-image-based color encodings.

PhotoYCC space is based on color appearance defined in terms of colorimetry and a specified set of encoding reference viewing conditions. The conditions are consistent with those of outdoor scenes. Encoded colorimetric values are expressed according to the properties of a specified reference image-capture device. This hypothetical device produces color values representative of the colorimetry of original scenes. The device is defined in terms of a set of reference primaries and an optoelectronic transfer characteristic based on Recommendations ITU-R BT.709 and ITU-R BT.601. However, unlike the electronic cameras described by those recommendations, the reference image-capture device has an extended luminance dynamic range, and its color gamut is not restricted to that defined by positive intensities of its reference RGB primaries.

Digital Color Management: Encoding Solutions, Second Edition Edward J. Giorgianni and Thomas E. Madden
 ISBN: 978-0-470-51244-9

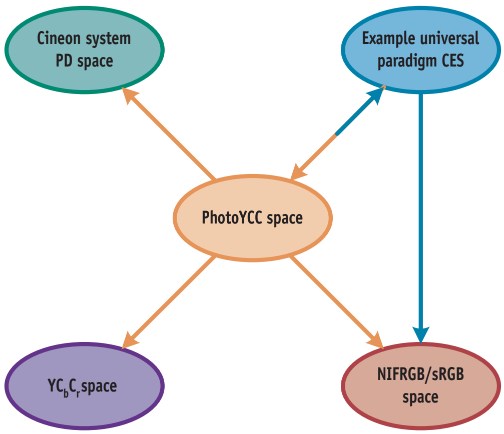


Figure G.1 The interchange transformations described in this appendix.

Example Unified Paradigm CES: The CES (Color Encoding Specification) used for most of the example color-management systems described in Part IV is based on color appearance defined in terms of colorimetry and a specified set of encoding reference viewing conditions. Those conditions are consistent with indoor viewing of reflection images.

YC_bC_r space: YC_bC_r values are used in many software applications. Like PhotoYCC space values, these values are based on the properties of a reference electronic camera defined according to Recommendations ITU-R BT.709 and ITU-R BT.601. However, with YC_bC_r the upper limit of the encoded luminance dynamic range is limited to the luminance factor of a perfect white. In addition, the encoded color gamut is restricted to that defined by positive intensities of the reference RGB primaries.

Cineon system PD space: The color encoding of the Cineon digital film system is based on printing-density measurements from scanned motion picture negative films. Printing-density values are measured according to a specified set of reference spectral responsivities defined by the spectral sensitivities of a reference print medium and the spectral properties of a reference printer light source.

NIFRGB/sRGB space: NIFRGB values were used in the FlashPix™ format, which was defined in a specification and a test suite developed and published by Eastman Kodak Company in collaboration with Hewlett-Packard Company, Live Picture Inc., and Microsoft Corporation. NIFRGB values, and the identical sRGB values, are defined according to the properties of a reference CRT-based monitor viewed

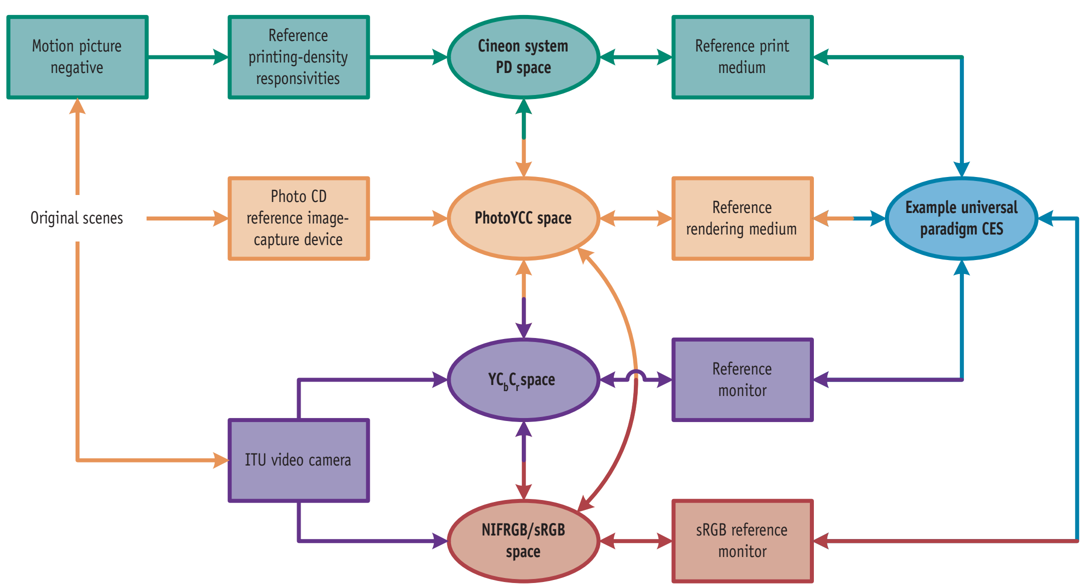


Figure G.2 Relationships of the described color encoding specifications to original scenes and reproduced images.

in a specified set of encoding reference viewing conditions. The upper limit of the encodable luminance dynamic range is restricted to the luminance factor of the maximum white stimulus that can be produced on that monitor. The encodable color gamut is restricted to that defined by positive intensities of the reference monitor primaries.

All of these CESs can be linked, as indicated by the arrows of Figures G.1 and G.2. In some cases, however, information will be lost in the transformation of color-image data from one encoding to another. Details of several of the more important transformations follow.

A method for transforming PhotoYCC space values to example Unified Paradigm CES values

Both the Kodak PhotoYCC color interchange space and the CES used for the example color-management systems described in Part IV can support general color interchange. Each uses a CES that is fully defined in terms of colorimetry, data metric, and associated viewing conditions. Each provides an unambiguous specification of color, and each can represent any given color, regardless of its origin. The properties of the two specifications are compared in Table G.1.

When appropriate procedures such as those described here are used, color interchange of PhotoYCC space values to and from the example CES values is unrestricted. The interchange transformation determines code values, in terms of the respective data metrics, that represent equivalent color stimuli in the respective viewing conditions. The determination of this equivalence includes consideration of the properties of actual imaging media and devices.

The encoding reference viewing conditions of the scene-based Kodak PhotoYCC color interchange space are typical of outdoor environments, whereas the encoding reference viewing conditions of the example CES are typical of indoor environments. A transform between the two must account for these different viewing conditions. In addition, the transform must account for any differences in the methods used in making colorimetric measurements and for any differences in the data metrics of the two CESs. These factors are compared and discussed below:

- ***Colorimetric measurements:*** Both CESs are derived from flareless (or flare-corrected) colorimetric measurements based on the CIE 1931 Standard Colorimetric Observer. Therefore, no adjustments for measurement differences are required in the color interchange transformation.

Table G.1 Comparison of PhotoYCC specifications and the CES (Color Encoding Specification) of the example color-management system.

	Photo YCC	Example CES
Colorimetry	1931 CIE flareless measurement	1931 CIE flareless measurement
Data metric	PhotoYCC space	CIELAB
	$Y = (255/1.402) Luma$	$CV_1 = 2.10L^*$
	$C_1 = 111.40\ Chroma_1 + 156$	$CV_2 = a^* + 128$
	$C_2 = 135.64\ Chroma_2 + 137$	$CV_3 = b^* + 128$
Viewing conditions:		
Viewing flare	None	1.0%
Image surround	Average	average
Luminance level	>1 600 cd/m^2	60–160 cd/m^2
Adaptive white	$x = 0.3127$, $y = 0.3290$	$x = 0.3457$, $y = 0.3585$

- ***Viewing flare:*** The transformation must include compensations for the physical reductions in both luminance contrast and color saturation associated with the higher level of viewing flare of the encoding reference viewing environment specified for the example CES.
- ***Surround type:*** Both reference viewing environments assume a normal (or average) image surround; therefore, no adjustments for observer lateral-brightness adaptation or general-brightness adaptation are required in the color-interchange transformation.
- ***Luminance level:*** Compensations must be included for the perceived reductions in luminance contrast and colorfulness associated with the lower luminance levels of the example CES reference viewing environment.
- ***Observer adaptive white:*** Colorimetric adjustments must be included to account for the differences in the respective observer's adaptive white chromaticity coordinates specified for the PhotoYCC space and example CES encoding reference viewing environments.
- ***Reference primaries:*** Colorimetric adjustments must be included to account for the differences in the reference primaries of the PhotoYCC space and example CES data metrics.
- ***Numerical units:*** The color-interchange transformation must include appropriate mathematical conversions to account for numerical differences between the PhotoYCC space and example CES data metrics.

The steps involved in the transformation of PhotoYCC space values to example CES values are shown in Figure G.3 and include the following:

- A data metric conversion from PhotoYCC space Y, C_1, C_2 digital code values to CIE XYZ tristimulus values.
- Adjustment of the XYZ values to account for the change in adaptive white chromaticity coordinates from those of D_{65} to those of D_{50}.
- Further adjustment of the XYZ values to compensate for the lower luminance levels of the example CES reference viewing conditions.
- Further adjustment of the XYZ values to compensate for the greater level of viewing flare in the example CES reference viewing conditions.
- A metric conversion of the adjusted XYZ values to CIELAB L^*, a^*, b^* values.
- Scaling of the CIELAB L^*, a^*, b^* values to form digital code values, CV_1, CV_2, and CV_3, as defined by the data metric of the example CES.

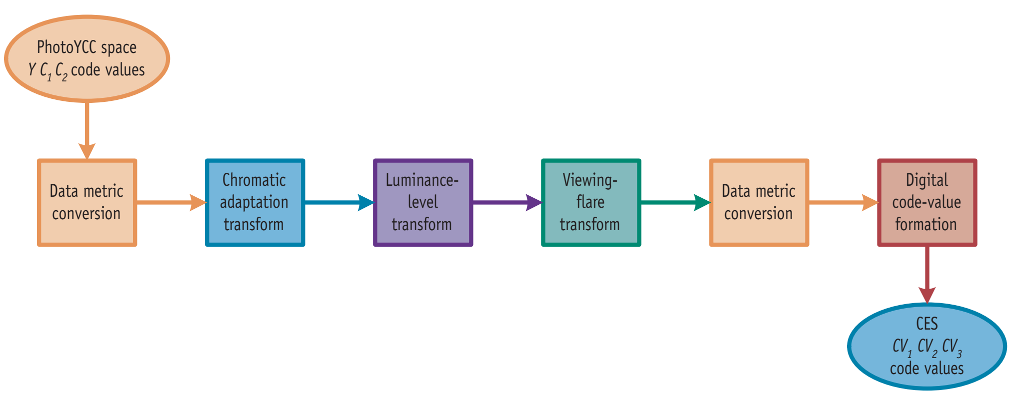


Figure G.3 Transformation from PhotoYCC space to the color encoding specification (CES) of the example color-management systems described in Part IV.

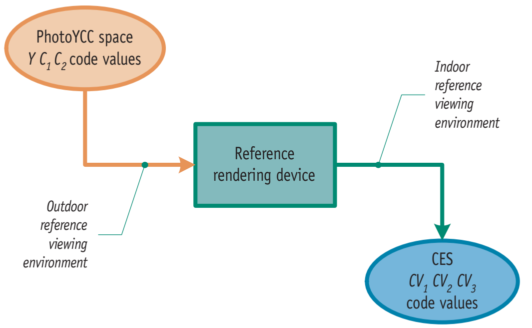


Figure G.4 Use of a reference rendering device in the transformation of PhotoYCC space values to example CES values.

The following transformation from PhotoYCC space values to example CES values represents one method that can be used to account for the differences in the reference viewing environments and numerical encoding methods of the two CESs. This transformation method includes the use of a reference rendering device, as shown in Figure G.4.

The reference device provides a means for converting values encoded with respect to a reference environment normally associated with the viewing of outdoor scenes, to values encoded with respect to a reference environment normally associated with the viewing of reproduced images, in a way that is consistent with the properties of actual imaging media and devices. Inclusion of the reference rendering device in the transformation ensures that image data transformed by this procedure will be compatible with image data of reproduced images that have been encoded directly in terms of example CES values.

The reference rendering device for this transformation is defined to be a hypothetical additive color device having the following characteristics:

- The capability of producing, in the reference viewing environment, a white of $L^* \approx 121$ (an optical density of approximately –0.22) at the chromaticity of CIE Standard Illuminant D_{50}.
- The capability of producing, in the absence of viewing flare, a black of $L^* \approx 0$ (an optical density of approximately 4.0) at the chromaticity of CIE Standard Illuminant D_{50}.

Table G.2 CIE x, y, z and u', v' chromaticity coordinates for the RGB primaries of the example CES reference rendering device and for CIE Standard Illuminant D_{50}.

	Red	Green	Blue	D_{50}
x	0.8730	0.1750	0.0850	0.3457
y	0.1440	0.9270	0.0001	0.3585
z	−0.0170	−0.1020	0.9149	0.2958
u'	1.1710	0.0508	0.1201	0.2092
v'	0.4346	0.6057	0.0003	0.4881

- No crosstalk among its color channels, i.e., red output is affected only by red input, green output only by green input, and blue output only by blue input.
- Red, green, and blue primaries defined by the CIE chromaticity values given above in Table G.2.
- An equal *RGB* rendering characteristic, as shown in Figure G.5. Sample input/output values for this characteristic are listed in Table G.3.

In the first step of the transformation, luma and chroma values are computed from PhotoYCC space Y, C_1, and C_2 digital values. For 24-bit (8 bits per channel) encoding, luma and chroma values are computed according to the following equations:

$$\begin{aligned} Luma &= \frac{1.402}{255} Y \\ Chroma_1 &= \frac{(C_1 - 156)}{111.40} \\ Chroma_2 &= \frac{(C_2 - 137)}{135.64} \end{aligned} \tag{G.1}$$

The resulting *Luma*, $Chroma_1$, and $Chroma_2$ values are converted to nonlinear values, $R'G'B'_{709}$, using the following matrix transformation:

$$\begin{bmatrix} R'_{709} \\ G'_{709} \\ B'_{709} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 1 \\ 1 & -0.194 & -0.509 \\ 1 & 1 & 0 \end{bmatrix} \begin{bmatrix} Luma \\ Chroma_1 \\ Chroma_2 \end{bmatrix} \tag{G.2}$$

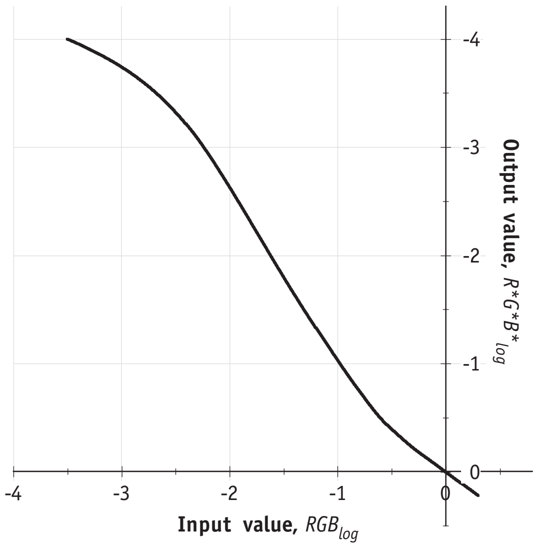


Figure G.5 Equal *RGB* characteristic for the example CES reference rendering device. The rendering characteristic incorporates compensations for viewing flare and for the lower luminances of the CES encoding reference viewing conditions.

The resulting $R'G'B'_{709}$ nonlinear values next are converted to linear values, RGB_{709}. This conversion to a linear space is necessary because the *RGB* values will subsequently be transformed to CIE *XYZ* values. That transformation then will allow the application of a chromatic adaptation transform to the *XYZ* values to account for the change in adaptive white chromaticity coordinates from those of D_{65} to those of D_{50}.

The $R'G'B'_{709}$ nonlinear values are converted to linear values, RGB_{709}, using the equations shown below.

For $R'G'B'_{709} \geq 0.081$:

$$R_{709} = \left(\frac{R'_{709} + 0.099}{1.099}\right)^{(1/0.45)}$$
$$G_{709} = \left(\frac{G'_{709} + 0.099}{1.099}\right)^{(1/0.45)} \qquad \text{(G.3a)}$$
$$B_{709} = \left(\frac{B'_{709} + 0.099}{1.099}\right)^{(1/0.45)}$$

Table G.3 Equal *RGB* characteristic for the example CES reference rendering device, as shown in Figure G.5.

RGB_{log} Input Value	$R^*\ G^*\ B^*{}_{log}$ Output Value
≤ -3.50	-4.00
-3.45	-3.98
-3.30	-3.91
-3.15	-3.83
-3.00	-3.74
-2.85	-3.64
-2.70	-3.52
-2.55	-3.38
-2.40	-3.21
-2.25	-3.01
-2.10	-2.79
-1.95	-2.55
-1.80	-2.30
-1.65	-2.05
-1.50	-1.80
-1.35	-1.56
-1.20	-1.33
-1.05	-1.10
-0.90	-0.88
-0.75	-0.68
-0.60	-0.49
-0.45	-0.33
-0.30	-0.21
-0.15	-0.09
$+0.00$	$+0.00$
$+0.15$	$+0.11$
$\geq +0.30$	$+0.22$

For $R'G'B'_{709} \leq -0.081$:

$$R_{709} = -\left(\frac{R'_{709} - 0.099}{-1.099}\right)^{(1/0.45)}$$
$$G_{709} = -\left(\frac{G'_{709} - 0.099}{-1.099}\right)^{(1/0.45)} \qquad \text{(G.3b)}$$
$$B_{709} = -\left(\frac{B'_{709} - 0.099}{-1.099}\right)^{(1/0.45)}$$

For $-0.081 < R'G'B'_{709} < 0.081$:

$$R_{709} = \frac{R'_{709}}{4.5}$$
$$G_{709} = \frac{G'_{709}}{4.5} \qquad \text{(G.3c)}$$
$$B_{709} = \frac{B'_{709}}{4.5}$$

The resulting RGB_{709} linear values are converted to CIE $XYZ_{D_{65}}$ values using the following matrix transformation:

$$\begin{bmatrix} X_{D_{65}} \\ Y_{D_{65}} \\ Z_{D_{65}} \end{bmatrix} = \begin{bmatrix} 41.24 & 35.76 & 18.05 \\ 21.26 & 71.52 & 7.22 \\ 1.93 & 11.92 & 95.05 \end{bmatrix} \begin{bmatrix} R_{709} \\ G_{709} \\ B_{709} \end{bmatrix} \qquad \text{(G.4)}$$

The resulting CIE $XYZ_{D_{65}}$ values then are transformed to account for the different adaptive whites of the two color specifications. The following von Kries chromatic adaptation matrix is used for this purpose:

$$\begin{bmatrix} X_{D_{50}} \\ Y_{D_{50}} \\ Z_{D_{50}} \end{bmatrix} = \begin{bmatrix} 1.0161 & 0.0553 & -0.0522 \\ 0.0060 & 0.9956 & -0.0012 \\ 0.0000 & 0.0000 & 0.7576 \end{bmatrix} \begin{bmatrix} X_{D_{65}} \\ Y_{D_{65}} \\ Z_{D_{65}} \end{bmatrix} \qquad \text{(G.5)}$$

Transformation to the example CES must include compensation for both the physical and the perceptual reductions in luminance contrast and color saturation that result from the higher viewing flare and lower luminance levels of the example CES reference viewing conditions. This is accomplished by using the previously described reference rendering device.

In this procedure, which will be detailed on the following pages, $XYZ_{D_{50}}$ values are converted to corresponding values expressed in terms of the red, green, and blue primaries of the reference rendering device. A nonlinear mapping corresponding to the equal RGB grayscale of the reference rendering device then is applied, and the mapped values are converted to D_{50} tristimulus values, XYZ_{CES}. In addition to providing the required luminance contrast and color-saturation compensations, these operations also have the desirable effect of mapping color values into the gamut of the reference rendering device. This results in transformed color values that are consistent with the properties of practical imaging media.

In the first step of this procedure, the CIE $XYZ_{D_{50}}$ values from Equation (G.5) are converted to RGB tristimulus values for the red, green, and blue primaries of the reference device, as defined in Table G.2, using the following matrix transformation:

$$\begin{bmatrix} R \\ G \\ B \end{bmatrix} = \begin{bmatrix} 1.4521 & -0.2890 & -0.1349 \\ -0.1929 & 1.1713 & 0.0178 \\ 0.0015 & 0.1173 & 1.0682 \end{bmatrix} \begin{bmatrix} \dfrac{X_{D_{50}}}{100} \\ \dfrac{Y_{D_{50}}}{100} \\ \dfrac{Z_{D_{50}}}{100} \end{bmatrix} \qquad \text{(G.6)}$$

The logarithms of the RGB tristimulus values then are calculated:

$$R_{log} = \log_{10}(R)$$
$$G_{log} = \log_{10}(G) \qquad \text{(G.7)}$$
$$B_{log} = \log_{10}(B)$$

The resulting RGB_{log} values are transformed to $R^*G^*B^*_{log}$ values using the lookup table listed in Table G.3 and shown in Figure G.5. The resulting $R^*G^*B^*_{log}$ values are converted to linear values,

$$R^*_{log} = G3\left(R_{log}\right)$$
$$G^*_{log} = G3\left(G_{log}\right) \qquad \text{(G.8a)}$$
$$B^*_{log} = G3\left(B_{log}\right)$$

$$R^* = 10^{R^*_{log}}$$
$$G^* = 10^{G^*_{log}} \qquad \text{(G.8b)}$$
$$B^* = 10^{B^*_{log}}$$

and transformed to XYZ_{CES} tristimulus values using the following matrix transformation:

$$\begin{bmatrix} X_{CES} \\ Y_{CES} \\ Z_{CES} \end{bmatrix} = \begin{bmatrix} 71.07 & 16.66 & 8.70 \\ 11.72 & 88.27 & 0.01 \\ -1.38 & -9.71 & 93.61 \end{bmatrix} \begin{bmatrix} R^* \\ G^* \\ B^* \end{bmatrix} \qquad \text{(G.9)}$$

The resulting XYZ_{CES} tristimulus values then are converted to CIELAB values, $L^*a^*b^*_{CES}$, using the equations previously described in Appendix A and given again below. In these calculations, the tristimulus values for the reference white are those of CIE D_{50} ($X_{n_{CES}} = 96.42$, $Y_{n_{CES}} = 100.00$, and $Z_{n_{CES}} = 82.49$).

$$L^*_{CES} = 116\left(\frac{Y_{CES}}{Y_{n_{CES}}}\right)^{\frac{1}{3}} - 16 \quad \text{for } \frac{Y_{CES}}{Y_{n_{CES}}} > 0.008\,856$$

$$L^*_{CES} = 903.3\left(\frac{Y_{CES}}{Y_{n_{CES}}}\right) \quad \text{for } \frac{Y_{CES}}{Y_{n_{CES}}} \leq 0.008\,856$$

and

$$\begin{aligned} a^*_{CES} &= 500\left[f\left(\frac{X_{CES}}{X_{n_{CES}}}\right) - f\left(\frac{Y_{CES}}{Y_{n_{CES}}}\right)\right] \\ b^*_{CES} &= 200\left[f\left(\frac{Y_{CES}}{Y_{n_{CES}}}\right) - f\left(\frac{Z_{CES}}{Z_{n_{CES}}}\right)\right] \end{aligned} \tag{G.10}$$

where

$$f\left(\frac{X_{CES}}{X_{n_{CES}}}\right) = \left(\frac{X_{CES}}{X_{n_{CES}}}\right)^{\frac{1}{3}} \quad \text{for } \frac{X_{CES}}{X_{n_{CES}}} > 0.008\,856$$

$$f\left(\frac{X_{CES}}{X_{n_{CES}}}\right) = 7.787\left(\frac{X_{CES}}{X_{n_{CES}}}\right) + \frac{16}{116} \quad \text{for } \frac{X_{CES}}{X_{n_{CES}}} \leq 0.008\,856$$

$$f\left(\frac{Y_{CES}}{Y_{n_{CES}}}\right) = \left(\frac{Y_{CES}}{Y_{n_{CES}}}\right)^{\frac{1}{3}} \quad \text{for } \frac{Y_{CES}}{Y_{n_{CES}}} > 0.008\,856$$

$$f\left(\frac{Y_{CES}}{Y_{n_{CES}}}\right) = 7.787\left(\frac{Y_{CES}}{Y_{n_{CES}}}\right) + \frac{16}{116} \quad \text{for } \frac{Y_{CES}}{Y_{n_{CES}}} \leq 0.008\,856$$

$$f\left(\frac{Z_{CES}}{Z_{n_{CES}}}\right) = \left(\frac{Z_{CES}}{Z_{n_{CES}}}\right)^{\frac{1}{3}} \quad \text{for } \frac{Z_{CES}}{Z_{n_{CES}}} > 0.008\,856$$

$$f\left(\frac{Z_{CES}}{Z_{n_{CES}}}\right) = 7.787\left(\frac{Z_{CES}}{Z_{n_{CES}}}\right) + \frac{16}{116} \quad \text{for } \frac{Z_{CES}}{Z_{n_{CES}}} \leq 0.008\,856$$

Finally, the resulting $L^*a^*b^*_{CES}$ values are converted to digital code values, CV_1, CV_2, and CV_3, according to the 24-bit (8 bits per channel) data metric defined for the example CES. The CES digital code values are the nearest integers to the values determined from the following equations:

$$CV_1 = 2.10L^*_{CES} \tag{G.11a}$$

$$CV_2 = a^*_{CES} + 128 \tag{G.11b}$$

$$CV_3 = b^*_{CES} + 128 \tag{G.11c}$$

For implementation purposes, any sequence of consecutive matrix operations in the preceding transformation can be combined into a single matrix. Likewise, any sequence of consecutive one-dimensional functions can be combined into a single one-dimensional lookup table. The entire transformation also can be implemented in the form of a three-dimensional lookup table.

It should be noted that the use of the rendering characteristic shown in Figure G.5 results in L^*_{CES} values greater than 100. Such values are necessary in order to optimally encode the appearance of original scenes and to accurately encode the color appearance of high-quality reproductions, as viewed under optimal conditions. Currently, however, many CESs do not provide for L^* values greater than 100.

Such values can be avoided by the use of a rendering characteristic such as that shown in Figure G.6. However, use of such characteristics results in some compression of highlight detail, which may be undesirable when images are to be displayed on high-quality output devices and media. The scale factor of Equation (G.11a) should be modified to be consistent with any changes in the rendering characteristic. For example, if the rendering characteristic is modified such that L^* values are limited to 100, the scale factor should be changed from a value of 2.10 to a value of 2.55.

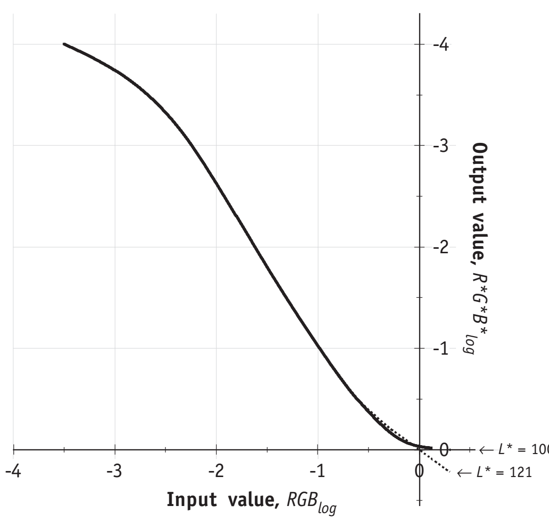


Figure G.6 Modified equal *RGB* characteristic for a CES reference rendering device with an L^* limit of 100 (the solid line) compared to a preferred characteristic having an L^* limit of approximately 121 (the dotted line).

A method for transforming example CES values to PhotoYCC space values

The transformation of example CES values to corresponding PhotoYCC color interchange space values is, of course, the reverse of that just shown. The first step in this transformation is a conversion of CES 8-bit code values to CIELAB $L^*a^*b^*_{CES}$ values:

$$\begin{aligned} L^*_{CES} &= \frac{CV_1}{2.10} \\ a^*_{CES} &= CV_2 - 128 \\ b^*_{CES} &= CV_3 - 128 \end{aligned} \tag{G.12}$$

The resulting CIELAB $L^*a^*b^*_{CES}$ values then are converted to CIE XYZ_{CES} values. In the following calculations, the tristimulus values for the reference white are those of CIE Standard Illuminant D_{50} ($X_{nCES} = 96.42$, $Y_{nCES} = 100.00$, and $Z_{nCES} = 82.49$).

$$Y_{CES} = Y_{nCES}\left(\frac{L^*_{CES}+16}{116}\right)^3 \quad \text{for } L^*_{CES} > 8$$

$$Y_{CES} = Y_{nCES}\left(\frac{L^*_{CES}}{903.3}\right) \quad \text{for } L^*_{CES} \le 8 \tag{G.13}$$

$$f\left(\frac{Y_{CES}}{Y_{nCES}}\right) = \left(\frac{Y_{CES}}{Y_{nCES}}\right)^{\frac{1}{3}} \quad \text{for } \frac{Y_{CES}}{Y_{nCES}} > 0.008\,856$$

$$f\left(\frac{Y_{CES}}{Y_{nCES}}\right) = 7.787\left(\frac{Y_{CES}}{Y_{nCES}}\right) + \frac{16}{116} \quad \text{for } \frac{Y_{CES}}{Y_{nCES}} < 0.008\,856$$

$$f\left(\frac{X_{CES}}{X_{nCES}}\right) = \left(\frac{a^*_{CES}}{500}\right) + f\left(\frac{Y_{CES}}{Y_{nCES}}\right)$$

$$f\left(\frac{Z_{CES}}{Z_{nCES}}\right) = f\left(\frac{Y_{CES}}{Y_{nCES}}\right) - \left(\frac{b^*_{CES}}{200}\right)$$

$$X_{CES} = X_{nCES}\left[f\left(\frac{X_{CES}}{X_{nCES}}\right)\right]^3 \quad \text{for } f\left(\frac{X_{CES}}{X_{nCES}}\right) > (0.008\,856)^{\frac{1}{3}}$$

$$X_{CES} = \frac{X_{nCES}}{7.787}\left[f\left(\frac{X_{CES}}{X_{nCES}}\right) - \frac{16}{116}\right] \quad \text{for } f\left(\frac{X_{CES}}{X_{nCES}}\right) \le (0.008\,856)^{\frac{1}{3}}$$

$$Z_{CES} = Z_{nCES}\left[f\left(\frac{Z_{CES}}{Z_{nCES}}\right)\right]^3 \quad \text{for } f\left(\frac{Z_{CES}}{Z_{nCES}}\right) > (0.008\,856)^{\frac{1}{3}}$$

$$Z_{CES} = \frac{Z_{nCES}}{7.787}\left[f\left(\frac{Z_{CES}}{Z_{nCES}}\right) - \frac{16}{116}\right] \quad \text{for } f\left(\frac{Z_{CES}}{Z_{nCES}}\right) \le (0.008\,856)^{\frac{1}{3}}$$

The next step in the transformation is to convert the resulting XYZ_{CES} tristimulus values to $R^*G^*B^*$ tristimulus values for the specified red, green, and blue primaries of the CES reference device, which were defined in Table G.2. This conversion is accomplished using the following matrix transformation:

$$\begin{bmatrix} R^* \\ G^* \\ B^* \end{bmatrix} = \begin{bmatrix} 1.4521 & -0.2890 & -0.1349 \\ -0.1929 & 1.1713 & 0.0178 \\ 0.0015 & 0.1173 & 1.0682 \end{bmatrix} \begin{bmatrix} \frac{X_{CES}}{100} \\ \frac{Y_{CES}}{100} \\ \frac{Z_{CES}}{100} \end{bmatrix} \tag{G.14}$$

The logarithms of the $R^*G^*B^*$ tristimulus values then are calculated:

$$\begin{aligned} R^*_{log} &= \log_{10}(R^*) \\ G^*_{log} &= \log_{10}(G^*) \\ B^*_{log} &= \log_{10}(B^*) \end{aligned} \tag{G.15}$$

The resulting $R^*G^*B^*_{log}$ values then are transformed to RGB_{log} values by going "backwards" through the reference rendering characteristic listed in Table G.3 and shown in Figure G.5:

$$\begin{aligned} R_{log} &= G3^{-1}\left(R^*_{log}\right) \\ G_{log} &= G3^{-1}\left(G^*_{log}\right) \\ B_{log} &= G3^{-1}\left(B^*_{log}\right) \end{aligned} \tag{G.16a}$$

The resulting RGB_{log} values then are converted to linear values, RGB:

$$\begin{aligned} R &= 10^{R_{log}} \\ G &= 10^{G_{log}} \\ B &= 10^{B_{log}} \end{aligned} \tag{G.16b}$$

The resulting RGB values next are transformed to $XYZ_{D_{50}}$ tristimulus values using the following matrix:

$$\begin{bmatrix} X_{D_{50}} \\ Y_{D_{50}} \\ Z_{D_{50}} \end{bmatrix} = \begin{bmatrix} 71.07 & 16.66 & 8.70 \\ 11.72 & 88.27 & 0.01 \\ -1.38 & -9.71 & 93.61 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix} \tag{G.17}$$

The $XYZ_{D_{50}}$ tristimulus values are transformed to $XYZ_{D_{65}}$ tristimulus values using the following von Kries chromatic adaptation matrix transformation:

$$\begin{bmatrix} X_{D_{65}} \\ Y_{D_{65}} \\ Z_{D_{65}} \end{bmatrix} = \begin{bmatrix} 0.9845 & -0.0547 & 0.0677 \\ -0.0059 & 1.0047 & 0.0012 \\ 0.0000 & 0.0000 & 1.3200 \end{bmatrix} \begin{bmatrix} X_{D_{50}} \\ Y_{D_{50}} \\ Z_{D_{50}} \end{bmatrix} \tag{G.18}$$

and then transformed to reference image-capture device RGB_{709} tristimulus values:

$$\begin{bmatrix} R_{709} \\ G_{709} \\ B_{709} \end{bmatrix} = \begin{bmatrix} 3.2410 & -1.5374 & -0.4986 \\ -0.9692 & 1.8760 & 0.0416 \\ 0.0556 & -0.2040 & 1.0570 \end{bmatrix} \begin{bmatrix} \frac{X_{D_{65}}}{100} \\ \frac{Y_{D_{65}}}{100} \\ \frac{Z_{D_{65}}}{100} \end{bmatrix} \tag{G.19}$$

Reference image-capture device RGB_{709} tristimulus values next are transformed to nonlinear values, $R'G'B'_{709}$, as follows:

For $RGB_{709} \geq 0.018$:

$$\begin{aligned} R'_{709} &= 1.099 R_{709}^{0.45} - 0.099 \\ G'_{709} &= 1.099 G_{709}^{0.45} - 0.099 \\ B'_{709} &= 1.099 B_{709}^{0.45} - 0.099 \end{aligned} \tag{G.20a}$$

For $RGB_{709} \leq -0.018$:

$$\begin{aligned} R'_{709} &= -1.099\,|R_{709}|^{0.45} + 0.099 \\ G'_{709} &= -1.099\,|G_{709}|^{0.45} + 0.099 \\ B'_{709} &= -1.099\,|B_{709}|^{0.45} + 0.099 \end{aligned} \tag{G.20b}$$

For $-0.018 < RGB_{709} < 0.018$:

$$\begin{aligned} R'_{709} &= 4.5 R_{709} \\ G'_{709} &= 4.5 G_{709} \\ B'_{709} &= 4.5 B_{709} \end{aligned} \tag{G.20c}$$

The $R'G'B'_{709}$ nonlinear values are transformed to luma and chroma values as follows:

$$\begin{bmatrix} Luma \\ Chroma_1 \\ Chroma_2 \end{bmatrix} = \begin{bmatrix} 0.299 & 0.587 & 0.114 \\ -0.299 & -0.587 & 0.886 \\ 0.701 & -0.587 & -0.114 \end{bmatrix} \begin{bmatrix} R'_{709} \\ G'_{709} \\ B'_{709} \end{bmatrix} \tag{G.21}$$

Finally, luma and chroma values are converted to digital code values. For 24-bit (8 bits per channel) encoding, PhotoYCC space values are formed according to the following equations:

$$\begin{aligned} Y &= \frac{255}{1.402} Luma \\ C_1 &= 111.40 Chroma_1 + 156 \\ C_2 &= 135.64 Chroma_1 + 137 \end{aligned} \tag{G.22}$$

A method for transforming PhotoYCC space values to YC_bC_r values

Kodak PhotoYCC color interchange space values can be transformed readily to YC_bC_r digital code values, which are used in many software applications. One commonly used definition for YC_bC_r is as follows:

$$\begin{aligned} Y_{YC_bC_r} &= 255E'_Y \\ C_b &= 255E'_{C_b} + 128 \\ C_r &= 255E'_{C_r} + 128 \end{aligned} \tag{G.23}$$

where E'_Y, E'_{C_b}, and E'_{C_r} are defined according to Recommendation ITU-R BT.601 (Rec. 601), and YC_bC_r digital code values are the nearest integers to the values computed from the preceding equations.

In Rec. 601, E'_Y, E'_{C_b}, and E'_{C_r} are derived from nonlinear exposure signals, E'_r, E'_g, and E'_b, as shown below, using the primaries and nonlinear equations of Recommendation ITU-R BT.709 (Rec. 709):

$$\begin{aligned} E'_Y &= 0.299E'_r + 0.587E'_g + 0.114E'_b \\ E'_{C_b} &= 0.564\left(E'_b - E'_Y\right) \\ E'_{C_r} &= 0.713\left(E'_r - E'_Y\right) \end{aligned} \tag{G.24}$$

Kodak PhotoYCC color interchange space also is based on Rec. 709 primaries and nonlinear equations. Therefore E'_Y, E'_{C_b}, and E'_{C_r} have the same basic definitions as PhotoYCC space Y, C_1, and C_2 values, although the scalings used in the respective data metrics are different, as shown below.

From Equations (G.23) and (G.24):

$$\begin{aligned} Y_{YC_bC_r} &= 255E'_Y \\ C_b &= 143.820\left(E'_b - E'_Y\right) + 128 \\ C_r &= 181.815\left(E'_r - E'_Y\right) + 128 \end{aligned} \tag{G.25}$$

The comparable PhotoYCC space equations are

$$\begin{aligned} Y_{YCC} &= \frac{255}{1.402} E'_Y \\ C_1 &= 111.40 \left(E'_b - E'_Y \right) + 156 \\ C_2 &= 135.64 \left(E'_r - E'_Y \right) + 137 \end{aligned} \qquad \text{(G.26)}$$

Therefore, Kodak PhotoYCC color interchange space values can be transformed to YC_bC_r values as follows:

$$\begin{aligned} Y_{YC_bC_r} &= 1.402 Y_{YCC} \\ C_b &= 1.291 C_1 - 73.400 \\ C_r &= 1.340 C_2 - 55.638 \end{aligned} \qquad \text{(G.27)}$$

It should be noted that in this transformation, PhotoYCC space highlight information above 100 % scene white will be clipped at code value 255. More satisfactory results generally are obtained by first performing an appropriate modification of the PhotoYCC space values. One method for doing this is to first convert PhotoYCC space Y, C_1, and C_2 values to $R'G'B'_{709}$ nonlinear values using Equations (G.1) and (G.2). The resulting $R'G'B'_{709}$ values then can be remapped, using a lookup table such as that shown in Figure G.7, so that the remapped values do not exceed 1.0. These remapped values then can be converted back to modified PhotoYCC space Y, C_1, and C_2 values using Equations (G.21) and (G.22), and to YC_bC_r values using Equations (G.27).

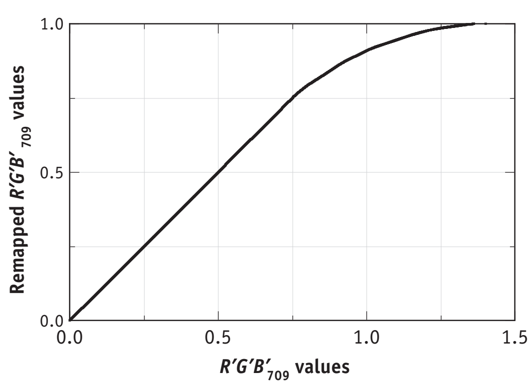


Figure G.7 Graphical representation of a lookup table for remapping PhotoYCC space $R'G'B'_{709}$ values. This remapping prevents clipping of highlight information in the transformation to YC_bC_r values.

YC_bC_r values can be transformed to PhotoYCC space values using the inverse of the procedure that has been described here. However, the result of attempting to expand the YC_bC_r encoded highlight information may not be satisfactory. It also should be noted that the color gamut of YC_bC_r is significantly smaller than that of Kodak PhotoYCC color interchange space. Therefore, the color quality of an image transformed to PhotoYCC space values from YC_bC_r values may not match that of an image encoded directly in terms of PhotoYCC space values.

A method for transforming PhotoYCC space values to Cineon system values

Cineon system code values correspond to printing densities for motion picture negative films. The printing densities are determined based on the spectral sensitivities of a reference print film and the spectral power distribution of the light source of a reference printer (Chapter 8, Appendix B). The basic steps involved in the transformation of PhotoYCC space YC_1C_2 values to Cineon system code values are shown in Figure G.8. These include the following operations:

- A data metric conversion from PhotoYCC space Y, C_1, and C_2 values to reference image-capture device R, G, and B tristimulus values.

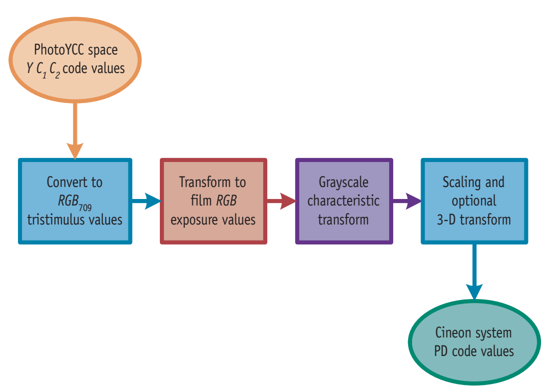


Figure G.8 Transformation of PhotoYCC space Y, C_1, and C_2 code values to Cineon system printing-density code values.

- A transformation of those R, G, and B tristimulus values to corresponding R, G, and B exposure values for a representative motion picture negative film.
- Mapping the resulting R, G, and B film exposure values through the grayscale characteristic of a representative motion picture negative film to form channel-independent printing-density values.
- Scaling the channel-independent printing-density values, and optionally applying a three-dimensional transform, to obtain Cineon system 10-bit code values.

In the first step of the transformation, luma and chroma values are computed from PhotoYCC space Y, C_1, and C_2 digital values. Luma and chroma values are computed from 24-bit (8 bits per channel) encoded PhotoYCC space values according to the following equations:

$$\begin{aligned} Luma &= \frac{1.402}{255}Y \\ Chroma_1 &= \frac{(C_1 - 156)}{111.40} \\ Chroma_2 &= \frac{(C_2 - 137)}{135.64} \end{aligned} \tag{G.28}$$

The resulting Luma, $Chroma_1$, and $Chroma_2$ values are converted to nonlinear values, $R'G'B'_{709}$, using the following matrix transformation:

$$\begin{bmatrix} R'_{709} \\ G'_{709} \\ B'_{709} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 1 \\ 1 & -0.194 & -0.509 \\ 1 & 1 & 0 \end{bmatrix} \begin{bmatrix} Luma \\ Chroma_1 \\ Chroma_2 \end{bmatrix} \tag{G.29}$$

The $R'G'B'_{709}$ nonlinear values are converted to linear tristimulus values, RGB_{709}, as shown below.

For $R'G'B'_{709} \geq 0.081$:

$$\begin{aligned} R_{709} &= \left(\frac{R'_{709} + 0.099}{1.099}\right)^{(1/0.45)} \\ G_{709} &= \left(\frac{G'_{709} + 0.099}{1.099}\right)^{(1/0.45)} \\ B_{709} &= \left(\frac{B'_{709} + 0.099}{1.099}\right)^{(1/0.45)} \end{aligned} \tag{G.30a}$$

For $R'G'B'_{709} \leq -0.081$:

$$\begin{aligned} R_{709} &= -\left(\frac{R'_{709} - 0.099}{-1.099}\right)^{(1/0.45)} \\ G_{709} &= -\left(\frac{G'_{709} - 0.099}{-1.099}\right)^{(1/0.45)} \\ B_{709} &= -\left(\frac{B'_{709} - 0.099}{-1.099}\right)^{(1/0.45)} \end{aligned} \tag{G.30b}$$

For $-0.081 < R'G'B'_{709} < 0.081$:

$$\begin{aligned} R_{709} &= \frac{R'_{709}}{4.5} \\ G_{709} &= \frac{G'_{709}}{4.5} \\ B_{709} &= \frac{B'_{709}}{4.5} \end{aligned} \tag{G.30c}$$

The resulting RGB_{709} tristimulus values next are converted to corresponding RGB_{film} exposure values. The matrix used in this conversion was determined such that, for a particular set of test colors, the differences between colorimetric values computed using the spectral sensitivities for a representative motion picture negative film and values converted from the RGB_{709} tristimulus values are minimized:

$$\begin{bmatrix} R_{film} \\ G_{film} \\ B_{film} \end{bmatrix} \begin{bmatrix} 0.8370 & 0.0800 & 0.0830 \\ 0.0023 & 0.9414 & 0.0563 \\ -0.0005 & 0.1046 & 0.8959 \end{bmatrix} \begin{bmatrix} R_{709} \\ G_{709} \\ B_{709} \end{bmatrix} \tag{G.31}$$

The logarithms of the RGB_{film} exposure values then are calculated:

$$\begin{aligned} R_{log} &= \log_{10}\left(R_{film}\right) \\ G_{log} &= \log_{10}\left(G_{film}\right) \\ B_{log} &= \log_{10}\left(B_{film}\right) \end{aligned} \tag{G.32}$$

The resulting RGB_{log} values are scaled and shifted as follows to form RGB values consistent with the input values of Table G.4:

$$\begin{aligned} R &= 1000R_{log} + 2500 \\ G &= 1000G_{log} + 2500 \\ B &= 1000B_{log} + 2500 \end{aligned} \tag{G.33}$$

The resulting RGB values can be transformed to Cineon-compatible channel-independent printing-density code values (10 bits per channel) using the lookup table defined in Table G.4. That table is based on the grayscale relationship of normalized

Table G.4 Cineon system code values (10 bits per channel) versus *R*, *G*, or *B* input value from Equation (G.33).

Input: *R*, *G*, or *B* value	Output: Cineon system code value	Input: *R*, *G*, or *B* value	Output: Cineon system code value
≤0	95	2100	560
100	95	2200	590
200	95	2300	623
300	97	2400	655
400	100	2500	685
500	104	2600	715
600	115	2700	745
700	137	2800	775
800	165	2900	805
900	195	3000	835
1000	225	3100	860
1100	255	3200	880
1200	285	3300	897
1300	318	3400	910
1400	350	3500	920
1500	380	3600	925
1600	410	3700	928
1700	440	3800	930
1800	470	3900	933
1900	500	≥4000	935
2000	530		

channel-independent printing density to relative log exposure factor for a representative motion picture film (Figure G.9).

While the results obtained using the grayscale table alone may be sufficient for many purposes, improved compatibility with actual film image data can be achieved if the channel-independent printing-density values obtained from the table are further transformed through an appropriate three-dimensional transform (matrix, polynomial, or 3-D LUT). As discussed in Chapters 12, 15, and 21, the three-dimensional transform accounts for various chemical and optical interactions associated with color photographic films. In most applications, a universal transform based on an average or representative film can be used. To more closely match the printing-density values of a particular film, both the grayscale table and the three-dimensional transform should be specific to that film.

A method for transforming example CES values to NIFRGB/sRGB code values

NIFRGB values, which were used in the FlashPix format, and the identical sRGB values specify color in terms of 8-bit *RGB* code values. The color represented by a set of NIFRGB/sRGB values is that which would be formed if those code values were input to a specified reference monitor and if the resulting color stimulus produced by the monitor were

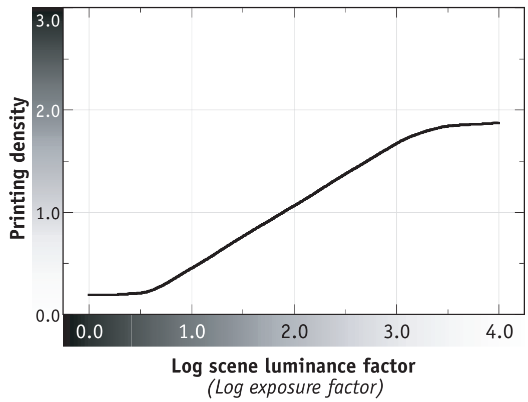


Figure G.9 Printing-density grayscale of a representative motion picture negative film. This grayscale is the basis for the values of Table G.4.

viewed according to a specified set of encoding reference viewing conditions (Table G.5).

The reference monitor is representative of those used on the majority of personal computers. It is defined in terms of a set of reference primaries and a characteristic curve relating nonlinear input signals to output relative intensities. The reference primaries are those defined in Recommendation ITU-R BT.709. The chromaticity coordinates of those primaries are given in Table G.6.

The grayscale characteristic for the reference monitor (Figure G.10) was designed to meet three criteria. First, because NIFRGB and sRGB code values are meant to be used directly as monitor code values, the characteristic had to be consistent with the characteristics of actual monitors. Second, the characteristic had to be such that its use produced images in which the visual effects of quantization were minimal. Third, for use in transformations, the characteristic had to have good mathematical reversibility.

Table G.5 NIFRGB/sRGB encoding reference viewing conditions.

Viewing flare	1.0 %
Image surround	average
Luminance level	80 cd/m^2
Adaptive white	$x = 0.3127$, $y = 0.3290$

Table G.6 Chromaticities of ITU-R BT.709 reference primaries and CIE Standard Illuminant D_{65}.

	Red	Green	Blue
x	0.6400	0.300	0.1500
y	0.3300	0.6000	0.0600
z	0.0300	0.1000	0.7900
u'	0.4507	0.1250	0.1754
v'	0.5229	0.5625	0.1579

The monitor characteristic is defined by the following equations relating nonlinear input signal values, $R'G'B'_{NIF}$, to output relative intensity values, RGB_{NIF}.

For $R'G'B'_{NIF} \geq 0.039\,29$:

$$\begin{aligned} R_{NIF} &= \left(\frac{R'_{NIF} + 0.055}{1.055}\right)^{2.4} \\ G_{NIF} &= \left(\frac{G'_{NIF} + 0.055}{1.055}\right)^{2.4} \\ B_{NIF} &= \left(\frac{B'_{NIF} + 0.055}{1.055}\right)^{2.4} \end{aligned} \tag{G.34a}$$

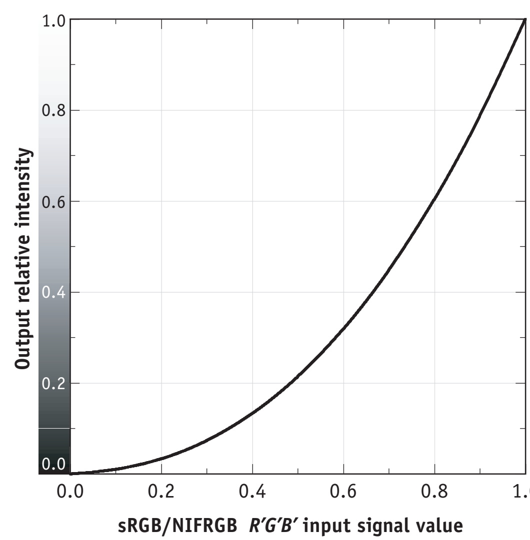


Figure G.10 Characteristic for the NIFRGB/sRGB reference monitor. The reference monitor is representative of those used on the majority of personal computers.

For $R'G'B'_{NIF} < 0.039\,29$:

$$R_{NIF} = \frac{R'_{NIF}}{12.92} \quad G_{NIF} = \frac{G'_{NIF}}{12.92} \quad B_{NIF} = \frac{B'_{NIF}}{12.92} \tag{G.34b}$$

The first step in the transformation of example CES values to visually equivalent NIFRGB/sRGB values is a conversion of CES code values to CIELAB $L^*a^*b^*_{CES}$ values:

$$L^*_{CES} = \frac{CV_1}{2.10} \quad a^*_{CES} = CV_2 - 128 \quad b^*_{CES} = CV_3 - 128 \tag{G.35}$$

The resulting CIELAB $L^*a^*b^*_{CES}$ values then are converted to CIE XYZ_{CES} values. In the following calculations, the tristimulus values for the reference white are those of CIE Standard Illuminant D_{50} ($X_{n_{CES}} = 96.42$, $Y_{n_{CES}} = 100.00$, and $Z_{n_{CES}} = 82.49$).

$$Y_{CES} = Y_{n_{CES}} \left(\frac{L^*_{CES} + 16}{116}\right)^3 \text{ for } L^*_{CES} > 8$$

$$Y_{CES} = Y_{n_{CES}} \left(\frac{L^*_{CES}}{903.3}\right) \text{ for } L^*_{CES} \le 8 \tag{G.36}$$

$$f\left(\frac{Y_{CES}}{Y_{n_{CES}}}\right) = \left(\frac{Y_{CES}}{Y_{n_{CES}}}\right)^{\frac{1}{3}} \text{ for } \frac{Y_{CES}}{Y_{n_{CES}}} > 0.008\,856$$

$$f\left(\frac{Y_{CES}}{Y_{n_{CES}}}\right) = 7.787\left(\frac{Y_{CES}}{Y_{n_{CES}}}\right) + \frac{16}{116} \text{ for } \frac{Y_{CES}}{Y_{n_{CES}}} \le 0.008\,856$$

$$f\left(\frac{X_{CES}}{X_{n_{CES}}}\right) = \left(\frac{a^*_{CES}}{500}\right) + f\left(\frac{Y_{CES}}{Y_{n_{CES}}}\right)$$

$$f\left(\frac{Z_{CES}}{Z_{n_{CES}}}\right) = f\left(\frac{Y_{CES}}{Y_{n_{CES}}}\right) - \left(\frac{b^*_{CES}}{200}\right)$$

$$X_{CES} = X_{n_{CES}} \left[f\left(\frac{X_{CES}}{X_{n_{CES}}}\right)\right]^3 \text{ for } f\left(\frac{X_{CES}}{X_{n_{CES}}}\right) > (0.008\,856)^{\frac{1}{3}}$$

$$X_{CES} = \frac{X_{n_{CES}}}{7.787} \left[f\left(\frac{X_{CES}}{X_{n_{CES}}}\right) - \frac{16}{116}\right] \text{ for } f\left(\frac{X_{CES}}{X_{n_{CES}}}\right) \le (0.008\,856)^{\frac{1}{3}}$$

$$Z_{CES} = Z_{n_{CES}} \left[f\left(\frac{Z_{CES}}{Z_{n_{CES}}}\right)\right]^3 \text{ for } f\left(\frac{Z_{CES}}{Z_{n_{CES}}}\right) > (0.008\,856)^{\frac{1}{3}}$$

$$Z_{CES} = \frac{Z_{n_{CES}}}{7.787} \left[f\left(\frac{Z_{CES}}{Z_{n_{CES}}}\right) - \frac{16}{116}\right] \text{ for } f\left(\frac{Z_{CES}}{Z_{n_{CES}}}\right) \le (0.008\,856)^{\frac{1}{3}}$$

The next step in the transformation is to convert the resulting XYZ_{CES} tristimulus values, which are based on D_{50} adaptive white chromaticity coordinates, to visually equivalent $XYZ_{D_{65}}$ tristimulus values, using the following von Kries chromatic adaptation matrix transformation:

$$\begin{bmatrix} X_{D_{65}} \\ Y_{D_{65}} \\ Z_{D_{65}} \end{bmatrix} = \begin{bmatrix} 0.9845 & -0.0547 & 0.0677 \\ -0.0059 & 1.0047 & 0.0012 \\ 0.0000 & 0.0000 & 1.3200 \end{bmatrix} \begin{bmatrix} X_{CES} \\ Y_{CES} \\ Z_{CES} \end{bmatrix} \tag{G.37}$$

The resulting $XYZ_{D_{65}}$ tristimulus values next are transformed to RGB_{NIF} tristimulus values:

$$\begin{bmatrix} R_{NIF} \\ G_{NIF} \\ B_{NIF} \end{bmatrix} = \begin{bmatrix} 3.2410 & -1.5374 & -0.4986 \\ -0.9692 & 1.8760 & 0.0416 \\ 0.0556 & -0.2040 & 1.0570 \end{bmatrix} \begin{bmatrix} \frac{X_{D_{65}}}{100} \\ \frac{Y_{D_{65}}}{100} \\ \frac{Z_{D_{65}}}{100} \end{bmatrix} \tag{G.38}$$

The resulting RGB_{NIF} tristimulus values, which represent reference monitor intensity values, are transformed to nonlinear values, $R'G'B'_{NIF}$, using equations that are the inverse of monitor characteristic given by Equations (G.34).

For $RGB_{NIF} \ge 0.003\,04$:

$$R'_{NIF} = 1.055R_{NIF}^{(1/2.4)} - 0.055 \quad G'_{NIF} = 1.055G_{NIF}^{(1/2.4)} - 0.055 \quad B'_{NIF} = 1.055B_{NIF}^{(1/2.4)} - 0.055 \tag{G.39}$$

For $RGB_{NIF} < 0.003\,04$:

$$R'_{NIF} = 12.92R_{NIF} \quad G'_{NIF} = 12.92G_{NIF} \quad B'_{NIF} = 12.92B_{NIF} \tag{G.40}$$

Finally, the resulting $R'G'B'_{NIF}$ values are converted to digital code values. For 24-bit (8 bits per channel) encoding, NIFRGB/sRGB values are the nearest integers to the values determined from the following equations:

$$R_{NIFRGB} = 255R'_{NIF} \quad G_{NIFRGB} = 255G'_{NIF} \quad B_{NIFRGB} = 255B'_{NIF} \tag{G.41}$$

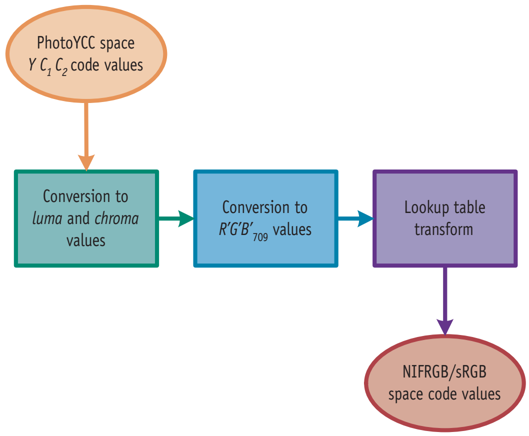


Figure G.11 Relationship of PhotoYCC space values to NIFRGB/sRGB values.

A method for transforming PhotoYCC space values to NIFRGB/sRGB code values

NIFRGB and sRGB values are closely related to Kodak PhotoYCC color interchange space values. The relationship is shown in Figure G.11.

In the first step of the transformation of PhotoYCC space values to NIFRGB/sRGB values, luma

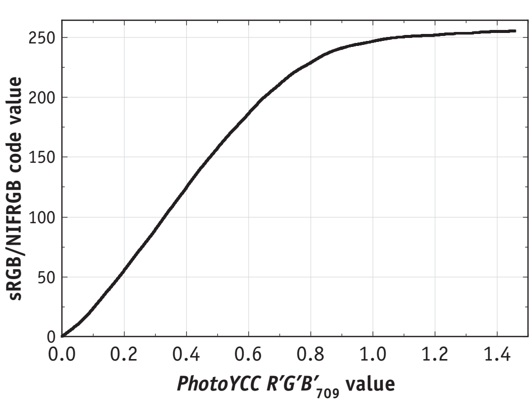


Figure G.12 Graphical representation of a lookup table for mapping PhotoYCC space $R'G'B'_{709}$ nonlinear values to NIFRGB/sRGB code values. The mapping prevents clipping of highlight information.

Table G.7 A lookup table for mapping $R'G'B'_{709}$ values to NIFRGB/sRGB code values.

Input Value $R'G'B'_{709}$	Output Value *NIFRGB/sRGB*
$\leq$ 0.00	0
0.05	10.15
0.10	23.44
0.15	39.17
0.20	55.23
0.25	72.35
0.30	89.13
0.35	107.06
0.40	124.35
0.45	141.63
0.50	157.24
0.55	172.15
0.60	186.07
0.65	199.58
0.70	210.65
0.75	221.07
0.80	228.59
0.85	236.00
0.90	240.79
0.95	243.94
1.00	246.41
1.05	248.92
1.10	250.17
1.15	251.04
1.20	251.64
1.25	252.88
1.30	253.07
1.35	254.03
1.40	254.61
$\geq$ 1.402	255.00

and chroma values are computed from PhotoYCC space Y, C_1, and C_2 digital code values. For 24-bit (8 bits per channel) encoding, the luma and chroma values are computed according to the following equations:

$$\begin{aligned} Luma &= \frac{1.402}{255} Y \\ Chroma_1 &= \frac{(C_1 - 156)}{111.40} \\ Chroma_2 &= \frac{(C_2 - 137)}{135.64} \end{aligned} \tag{G.42}$$

The resulting *Luma*, $Chroma_1$, and $Chroma_2$ values are converted to nonlinear values, $R'G'B'_{709}$:

$$\begin{bmatrix} R'_{709} \\ G'_{709} \\ B'_{709} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 1 \\ 1 & -0.194 & -0.509 \\ 1 & 1 & 0 \end{bmatrix} \begin{bmatrix} Luma \\ Chroma_1 \\ Chroma_2 \end{bmatrix} \tag{G.43}$$

The $R'G'B'_{709}$ nonlinear values are converted to NIFRGB/sRGB values using the lookup table shown in Figure G.12 and listed in Table G.7:

$$\begin{aligned} R_{NIFRGB} &= G7\left(R'_{709}\right) \\ G_{NIFRGB} &= G7\left(G'_{709}\right) \\ B_{NIFRGB} &= G7\left(B'_{709}\right) \end{aligned} \tag{G.44}$$

The conversion scales and remaps the $R'G'B'_{709}$ values such that high-luminance information is not clipped. Note, however, that any negative $R'G'B'_{709}$ values are clipped at zero. The color gamut of NIFRGB/sRGB encoding therefore is limited to that which can be produced on the reference monitor.

NIFRGB/sRGB values can be transformed to Kodak PhotoYCC color interchange space values using the inverse of this procedure. However, the result of attempting to expand highlight information encoded in terms of NIFRGB/sRGB values may not be satisfactory. It also should be noted that the color gamut of NIFRGB/sRGB space is significantly smaller than that of Kodak PhotoYCC color interchange space. Therefore, the color quality of an image transformed to PhotoYCC space values from NIFRGB/sRGB values may not match that of an image encoded directly in terms of PhotoYCC space values. For that reason, the FlashPix format supports the encoding of images in terms of Kodak PhotoYCC color interchange space values as well as in terms of NIFRGB/sRGB values.

H

Color-Primary Conversions

Procedures for deriving 3×3 matrices for converting color values expressed in terms of one set of primaries to colorimetrically equivalent color values expressed in terms of a different set of color primaries are described below. The procedures can be used to convert *RGB* color values to colorimetrically equivalent CIE *XYZ* color values, CIE *XYZ* color values to colorimetrically equivalent *RGB* color values, and *RGB* color values expressed in terms of a one set of red, green, and blue primaries to colorimetrically equivalent *RGB* color values expressed in terms of a second set of red, green, and blue primaries. Also described is a related procedure for determining the color-matching functions for a given set of red, green, and blue primaries.

Matrix notation

A 3×3 matrix is a shorthand expression for three simultaneous equations. For example, the matrix expression

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} m_{11} & m_{12} & m_{13} \\ m_{21} & m_{22} & m_{23} \\ m_{31} & m_{32} & m_{33} \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix} \tag{H.1}$$

is equivalent to the set of simultaneous equations

$$\begin{aligned} X &= m_{11}R + m_{12}G + m_{13}B \\ Y &= m_{21}R + m_{22}G + m_{23}B \\ Z &= m_{31}R + m_{32}G + m_{33}B \end{aligned} \tag{H.2}$$

RGB to CIE *XYZ*

A matrix for converting *RGB* color values expressed in terms of a set of red, green, and blue primaries to colorimetrically equivalent CIE *XYZ* color values expressed in terms of CIE X, Y, and Z primaries can be derived from the chromaticities of the red, green, and blue primaries and the chromaticity of their *white point*, i.e., the stimulus formed by unit-normalized amounts of each primary. As we will do here, such matrices are often referred to as *phosphor matrices* because they can be used to determine CIE *XYZ* tristimulus values from a set of CRT *RGB* intensity values. Because such matrices have been associated with that application for so long, the terminology is still commonly used even when applied to other types of additive displays that do not use phosphors.

The general relationship between normalized *RGB* values and colorimetrically equivalent *XYZ* tristimulus values is given by

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} X_r & X_g & X_b \\ Y_r & Y_g & Y_b \\ Z_r & Z_g & Z_b \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix} \tag{H.3}$$

which expands to

$$\begin{aligned} X &= X_rR + X_gG + X_bB \\ Y &= Y_rR + Y_gG + Y_bB \\ Z &= Z_rR + Z_gG + Z_bB \end{aligned} \tag{H.4}$$

Digital Color Management: Encoding Solutions, Second Edition Edward J. Giorgianni and Thomas E. Madden
 ISBN: 978-0-470-51244-9

where

R, G, and B are white-point-normalized *RGB* tristimulus values for a color stimulus produced by the additive combination a set of red, green, and blue color primaries;

X, Y, and Z are the corresponding CIE *XYZ* tristimulus values for the additive stimulus;

X_r, Y_r, and Z_r are the *XYZ* tristimulus values for the red primary;

X_g, Y_g, and Z_g are the *XYZ* tristimulus values for the green primary; and

X_b, Y_b, and Z_b are the *XYZ* tristimulus values for the red primary.

The coefficient values for the matrix in the above conversions can be computed based on the x, y chromaticity coordinates of the individual red, green, and blue color primaries and of the reference white stimulus produced when the full intensities (or another set of intensities defined as producing the reference white stimulus) are additively combined. Conventionally, the *RGB* primary amounts are normalized relative to the contribution required by each to produce the reference white stimulus. Therefore the reference white stimulus is produced when $R = G = B = 1$. Also by convention, the *XYZ* tristimulus values are normalized such that the Y tristimulus value of the reference white stimulus is equal to 100. Thus Y values are specified in terms of *percent luminance factor*.

In the first step in computing the conversion matrix, a matrix $\boldsymbol{C}$, composed of the *RGB* color-primary chromaticities is formed, incorporating the relationship $z = 1 - x - y$:

$$\boldsymbol{C} = \begin{bmatrix} x_r & y_r & z_r \\ x_g & y_g & z_g \\ x_b & y_b & z_b \end{bmatrix} \tag{H.5}$$

A vector $\boldsymbol{w}$ consisting of the *XYZ* tristimulus values for the additive white-point stimulus also is formed:

$$\boldsymbol{w} = \begin{bmatrix} X_w \\ Y_w \\ Z_w \end{bmatrix} \tag{H.6}$$

where

$$\begin{aligned} Y_w &= 100 \\ X_w &= \frac{x_w Y_w}{y_w} = 100\frac{x_w}{y_w} \\ Z_w &= \frac{z_w Y_w}{y_w} = 100\frac{z_w}{y_w} = 100\frac{1 - x_w - y_w}{y_w} \end{aligned} \tag{H.7}$$

Next, a vector $\boldsymbol{p}$ is computed from the matrix product of the white-point tristimulus vector and the transpose of the inverse primary-chromaticity matrix:

$$\boldsymbol{p} = \left[\boldsymbol{C}^{-1}\right]^{\mathrm{T}} \boldsymbol{w} \tag{H.8}$$

$$\begin{bmatrix} p_r \\ p_g \\ p_b \end{bmatrix} = \left[\begin{bmatrix} x_r & y_r & z_r \\ x_g & y_g & z_g \\ x_b & y_b & z_b \end{bmatrix}^{-1}\right]^{\mathrm{T}} \begin{bmatrix} X_w \\ Y_w \\ Z_w \end{bmatrix} \tag{H.9}$$

The phosphor matrix, $\boldsymbol{M}$, then is the product of the transpose primary-chromaticity matrix and a diagonal matrix formed from the $\boldsymbol{p}$-vector coefficients:

$$\boldsymbol{M} = \boldsymbol{C}^{\mathrm{T}} \mathrm{diag}(\boldsymbol{p})$$

$$\begin{bmatrix} X_r & X_g & X_b \\ Y_r & Y_g & Y_b \\ Z_r & Z_g & Z_b \end{bmatrix} = \begin{bmatrix} x_r & y_r & z_r \\ x_g & y_g & z_g \\ x_b & y_b & z_b \end{bmatrix}^{\mathrm{T}} \begin{bmatrix} p_r & 0 & 0 \\ 0 & p_g & 0 \\ 0 & 0 & p_b \end{bmatrix} \tag{H.10}$$

Worked example

In the following worked example, a phosphor matrix for converting *RGB* color values expressed in terms of a particular set of red, green, and blue primaries will be derived. Note that, for clarity, the results are given only to two decimal places.

In this example, the x, y chromaticity coordinates of the color primaries and D_{65} white point for the ITU-R BT.709 encoding are given by

Color	x	y
Red	0.6400	0.3300
Green	0.3000	0.6000
Blue	0.1500	0.0600
White	0.3127	0.3290

First, the z chromaticity coordinates are computed for each of the color primaries and white point:

$$
\begin{aligned}
z_r &= 1 - 0.64 - 0.33 = 0.03 \\
z_g &= 1 - 0.30 - 0.60 = 0.10 \\
z_b &= 1 - 0.15 - 0.06 = 0.79 \\
z_w &= 1 - 0.3127 - 0.3290 = 0.3583
\end{aligned}
\tag{H.11}
$$

The primary chromaticity matrix $\boldsymbol{C}$ is formed:

$$
\boldsymbol{C} = \begin{bmatrix} 0.64 & 0.33 & 0.03 \\ 0.30 & 0.60 & 0.10 \\ 0.15 & 0.06 & 0.79 \end{bmatrix}
\tag{H.12}
$$

The $\boldsymbol{w}$ white-point tristimulus value vector is formed:

$$
\begin{aligned}
Y_w &= 100 \\
X_w &= 100\frac{0.3127}{0.3290} \\
&= 95.05 \\
Z_w &= 100\frac{0.3583}{0.3290} \\
&= 108.91 \\
\boldsymbol{w} &= \begin{bmatrix} 95.05 \\ 100 \\ 108.91 \end{bmatrix}
\end{aligned}
\tag{H.13}
$$

Next, the vector $\boldsymbol{p}$ is computed by multiplying the white-point tristimulus vector by the transpose of the inverse primary-chromaticity matrix:

$$
\begin{aligned}
\begin{bmatrix} p_r \\ p_g \\ p_b \end{bmatrix} &= \left[\begin{bmatrix} 0.64 & 0.33 & 0.03 \\ 0.30 & 0.60 & 0.10 \\ 0.15 & 0.06 & 0.79 \end{bmatrix}^{-1}\right]^{\mathrm{T}} \begin{bmatrix} 95.05 \\ 100 \\ 108.91 \end{bmatrix} \\
&= \begin{bmatrix} 64.44 \\ 119.19 \\ 120.32 \end{bmatrix}
\end{aligned}
\tag{H.14}
$$

The phosphor matrix $\boldsymbol{M}$ then is computed from product of the transpose primary-chromaticity matrix and a diagonal matrix formed from the $\boldsymbol{p}$-vector coefficients:

$$
\begin{aligned}
\boldsymbol{M} &= \begin{bmatrix} 0.64 & 0.33 & 0.03 \\ 0.30 & 0.60 & 0.10 \\ 0.15 & 0.06 & 0.79 \end{bmatrix}^{\mathrm{T}} \operatorname{diag} \begin{bmatrix} 64.44 \\ 119.19 \\ 120.32 \end{bmatrix} \\
&= \begin{bmatrix} 0.64 & 0.30 & 0.15 \\ 0.33 & 0.60 & 0.06 \\ 0.03 & 0.10 & 0.79 \end{bmatrix} \begin{bmatrix} 64.44 & 0 & 0 \\ 0 & 119.19 & 0 \\ 0 & 0 & 120.32 \end{bmatrix} \\
&= \begin{bmatrix} 41.24 & 35.76 & 18.05 \\ 21.26 & 71.52 & 7.22 \\ 1.93 & 11.92 & 95.05 \end{bmatrix}
\end{aligned}
\tag{H.15}
$$

The CIE tristimulus values for the color formed by relative amounts of the RGB primaries (balanced for D_{65} white point) then are given by

$$
\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} 41.24 & 35.76 & 18.05 \\ 21.26 & 71.52 & 7.22 \\ 1.93 & 11.92 & 95.05 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix}
\tag{H.16a}
$$

or

$$
\begin{aligned}
X &= 41.24R + 35.76G + 18.05B \\
Y &= 21.26R + 71.52G + 7.22B \\
Z &= 1.93R + 11.82G + 95.05B
\end{aligned}
\tag{H.16b}
$$

CIE *XYZ* to *RGB*

A matrix for converting CIE XYZ color values expressed in terms of CIE X, Y, and Z primaries to colorimetrically equivalent RGB color values expressed in terms of a given set of red, green, and blue primaries can be derived simply by forming the inverse of the phosphor matrix for those RGB primaries. Procedures for inverting 3×3 matrices can be found in numerous texts.

Conversions between sets of *RGB* primaries

The principles above can be applied to compute 3×3 matrices for converting between different sets of RGB primaries. However, this requires that the individual red, green, and blue primaries of one set be expressed in terms of the chromaticity coordinates of the other set. This relationship generally is not known directly. While a derivation of that relationship is possible and is documented in other texts, many find the involved procedure somewhat difficult to follow.

As an alternative, we suggest the more straightforward method given below. The method is based on combining the relationships of the two sets of *RGB* primaries with their respective *XYZ* tristimulus values. The procedure yields the same result as the more rigorous *RGB* derivation.

The general form for the matrix for converting from one set of *RGB* primaries to another is given by

$$RGB_2 = [\boldsymbol{M}_{1\rightarrow 2}]\, RGB_1 \tag{H.17}$$

$$\begin{bmatrix} R_2 \\ G_2 \\ B_2 \end{bmatrix} = \begin{bmatrix} m_{11} & m_{12} & m_{13} \\ m_{21} & m_{22} & m_{23} \\ m_{31} & m_{32} & m_{33} \end{bmatrix} \begin{bmatrix} R_1 \\ G_1 \\ B_1 \end{bmatrix} \tag{H.18}$$

where

$$\begin{aligned} R_2 &= m_{11}R_1 + m_{12}G_1 + m_{13}B_1 \\ G_2 &= m_{21}R_1 + m_{22}G_1 + m_{23}B_1 \\ B_2 &= m_{31}R_1 + m_{32}G_1 + m_{33}B_1 \end{aligned} \tag{H.19}$$

The phosphor matrices for the two sets of *RGB* primaries are first computed using the method described above:

$$\boldsymbol{M}_1 = \begin{bmatrix} X_{1r} & Y_{1r} & Z_{1r} \\ X_{1g} & Y_{1g} & Z_{1g} \\ X_{1b} & Y_{1b} & Z_{1b} \end{bmatrix} \tag{H.20}$$

$$\boldsymbol{M}_2 = \begin{bmatrix} X_{2r} & Y_{2r} & Z_{2r} \\ X_{2g} & Y_{2g} & Z_{2g} \\ X_{2b} & Y_{2b} & Z_{2b} \end{bmatrix} \tag{H.21}$$

The matrix for transforming RGB_1 into colorimetrically equivalent RGB_2 values is given by the product

$$\begin{aligned} M_{1\rightarrow 2} &= \boldsymbol{M}_2^{-1}\boldsymbol{M}_1 \\ &= \begin{bmatrix} X_{2r} & Y_{2r} & Z_{2r} \\ X_{2g} & Y_{2g} & Z_{2g} \\ X_{2b} & Y_{2b} & Z_{2b} \end{bmatrix}^{-1} \begin{bmatrix} X_{1r} & Y_{1r} & Z_{1r} \\ X_{1g} & Y_{1g} & Z_{1g} \\ X_{1b} & Y_{1b} & Z_{1b} \end{bmatrix} \end{aligned} \tag{H.22}$$

In this process, matrix $\boldsymbol{M}_1$ first converts RGB_1 values to CIE *XYZ* tristimulus values, and those *XYZ* values then are converted to RGB_2 values by the inverse of matrix $\boldsymbol{M}_2$. This sequence of successive matrix operations is mathematically combined into a single composite matrix relationship, $\boldsymbol{M}_{12}$.

Worked example

Assume the RGB_1 primaries again correspond to the ITU-R BT.709 reference primaries and are balanced for a CIE Standard Illuminant D_{65} white point. Then matrix $\boldsymbol{M}_1$ is as before:

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} 41.24 & 35.76 & 18.05 \\ 21.26 & 71.52 & 7.22 \\ 1.93 & 11.92 & 95.05 \end{bmatrix} \begin{bmatrix} R_1 \\ G_1 \\ B_1 \end{bmatrix} \tag{H.23}$$

Assume the RGB_2 primaries correspond to the ERIMM primaries, and D_{65} white point:

Color	*x*	*y*
Red	0.7347	0.2653
Green	0.1596	0.8404
Blue	0.0366	0.0001
White	0.3127	0.3290

Using the methods described above, the $\boldsymbol{M}_2$ matrix relationship is

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} 77.21 & 13.69 & 4.14 \\ 27.88 & 72.11 & 1.13 \\ 0.00 & 0.00 & 108.91 \end{bmatrix} \begin{bmatrix} R_2 \\ G_2 \\ B_2 \end{bmatrix} \tag{H.24}$$

The matrix for converting RGB_1 values into corresponding RGB_2 values is then

$$\begin{aligned} \boldsymbol{M}_{1\rightarrow 2} &= \boldsymbol{M}_2^{-1}\boldsymbol{M}_1 \\ &= \begin{bmatrix} 77.21 & 13.69 & 4.14 \\ 27.88 & 72.11 & 1.13 \\ 0.00 & 0.00 & 108.91 \end{bmatrix}^{-1} \\ &\quad \times \begin{bmatrix} 41.24 & 35.76 & 18.05 \\ 21.26 & 71.52 & 7.22 \\ 1.93 & 11.92 & 95.05 \end{bmatrix} \\ &= \begin{bmatrix} 0.5162 & 0.3021 & 0.1817 \\ 0.0953 & 0.8750 & 0.0297 \\ 0.0178 & 0.1094 & 0.8728 \end{bmatrix} \end{aligned} \tag{H.25}$$

So,

$$\begin{bmatrix} R_2 \\ G_2 \\ B_2 \end{bmatrix} = \begin{bmatrix} 0.5162 & 0.3021 & 0.1817 \\ 0.0953 & 0.8750 & 0.0297 \\ 0.0178 & 0.1094 & 0.8728 \end{bmatrix} \begin{bmatrix} R_1 \\ G_1 \\ B_1 \end{bmatrix} \tag{H.26a}$$

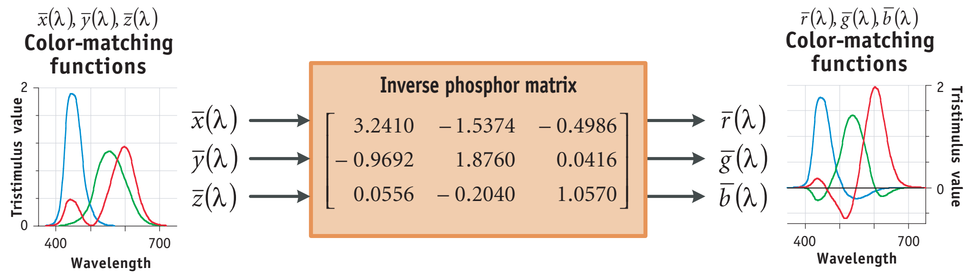


Figure H.1 The color-matching functions for a given set of red, green, and blue primaries can be determined using the inverse of the phosphor matrix ***M*** for the primaries together with the $\bar{x}(\lambda)$, $\bar{y}(\lambda)$, and $\bar{z}(\lambda)$ color-matching functions for the CIE Standard Colorimetric Observer.

or

$$\begin{aligned} R_2 &= 0.5162R_1 + 0.3021G_1 + 0.1817B_1 \\ G_2 &= 0.0953R_1 + 0.8750G_1 + 0.0297B_1 \\ B_2 &= 0.0178R_1 + 0.1094G_1 + 0.8728B_1 \end{aligned} \tag{H.26b}$$

Note the identity row sums in this conversion matrix. This is a necessary outcome because both sets of primaries initially were normalized to a common white point.

Color-matching functions for a set of primaries

One simple method for determining the color-matching functions for a given set of red, green, and blue primaries is to first derive the phosphor matrix ***M*** for the primaries, as described above. The inverse of that matrix then can be used with the $\bar{x}(\lambda)$, $\bar{y}(\lambda)$, and $\bar{z}(\lambda)$ color-matching functions for the CIE Standard Colorimetric Observer to determine the color-matching functions for the red, green, and blue primaries as shown in Equation (H27) and illustrated in Figure H.1:

$$cmf_{rgb} = \boldsymbol{M}^{-1} \cdot cmf_{xyz} \tag{H.27}$$

For example, the set of $\bar{r}(\lambda)$, $\bar{g}(\lambda)$, and $\bar{b}(\lambda)$ color-matching functions associated with the ITU-R BT.709 reference primaries, and balanced for a D_{65} white point, is computed as follows.

From the earlier example, the phosphor matrix ***M*** relating *RGB* values to XYZ values is as before:

$$\boldsymbol{M} = \begin{bmatrix} 41.24 & 35.76 & 18.05 \\ 21.26 & 71.52 & 7.22 \\ 1.93 & 11.92 & 95.05 \end{bmatrix} \tag{H.28}$$

The inverse matrix, normalized such that *R*, *G*, and *B* are all equal to 1.0 for a white object illuminated by CIE Standard Illuminant D_{65}, is

$$\begin{aligned} \boldsymbol{M}^{-1} &= \begin{bmatrix} 0.4124 & 0.3576 & 0.1805 \\ 0.2126 & 0.7152 & 0.0722 \\ 0.1930 & 0.1192 & 0.9505 \end{bmatrix}^{-1} \\ &= \begin{bmatrix} 3.2410 & -1.5374 & -0.4986 \\ -0.9692 & 1.8760 & 0.0416 \\ 0.0556 & -0.2040 & 1.0570 \end{bmatrix} \end{aligned} \tag{H.29}$$

Then,

$$\begin{bmatrix} \bar{r}(\lambda) \\ \bar{g}(\lambda) \\ \bar{b}(\lambda) \end{bmatrix} = \begin{bmatrix} 3.2410 & -1.5374 & -0.4986 \\ -0.9692 & 1.8760 & 0.0416 \\ 0.0556 & -0.2040 & 1.0570 \end{bmatrix} \times \begin{bmatrix} \bar{x}(\lambda) \\ \bar{y}(\lambda) \\ \bar{z}(\lambda) \end{bmatrix} \tag{H.30}$$

and for each wavelength, λ,

$$\begin{aligned} \bar{r}(\lambda) &= 3.2410 \cdot \bar{x}(\lambda) - 1.5374 \cdot \bar{y}(\lambda) \\ &\quad -0.4986 \cdot \bar{z}(\lambda) \\ \bar{g}(\lambda) &= -0.9692 \cdot \bar{x}(\lambda) + 1.8760 \cdot \bar{y}(\lambda) \\ &\quad + 0.0416 \cdot \bar{z}(\lambda) \\ \bar{b}(\lambda) &= 0.0556 \cdot \bar{x}(\lambda) - 0.2040 \cdot \bar{y}(\lambda) \\ &\quad + 1.0570 \cdot \bar{z}(\lambda) \end{aligned} \tag{H.31}$$

I

Mathematical Transforms

The color transformations described throughout this book can be implemented using a number of different mathematical techniques. The best method for a given application will depend on the hardware, software, and computational resources available. Nevertheless, different types of mathematical operations generally are appropriate for particular types of transformations, and the comments in this appendix regarding each are applicable in most circumstances.

One-dimensional lookup tables

A one-dimensional lookup table (1-D LUT) essentially is a list relating each individual input value to a corresponding output value. Such tables are an obvious choice for transforming a set of channel-independent image values to a different set of channel-independent values. Typical applications include reshaping grayscales for rendering or re-rendering, applying color-balance and exposure shifts, applying or removing camera nonlinearities in digital video and digital still images, and implementing calibration corrections for input and output devices. Table I.1 tabulates sample input/output pairs for an example 1-D LUT that might be used to prevent highlight clipping when extended luminance dynamic range input values are transformed to output 8-bit digital code values. The complete LUT is depicted graphically in Figure I.1.

In systems with limited resources, 1-D LUTs can be particularly useful. Consider, for example, a 24-bit color system (8 bits per color channel) employing 8-bit 1-D LUTs and a 3-D LUT (discussed below) limited in size to 64 cubed. In that system, it would be advantageous to perform any inherently 1-D transformations using the 1-D LUTs rather than relegate these transformations to the 3-D LUT. This allows a separation of signal processing into color operations, which must be performed in the 3-D LUT, and grayscale processing, which in this example is better performed using 1-D LUTs. Each 1-D LUT (one for each color channel) would have 256 explicit input and corresponding output data points, whereas the 3-D LUT would have only 64 data points (at the node locations along a 1-D slice of the table). Therefore a significant amount of interpolation of grayscale values would be required if a 3-D LUT alone were used to perform all of the signal processing. Our experience is that this type of interpolation often leads to problems, frequently in the form of hue shifts in neutrals and near-neutral colors.

Normalized 3 × 3 matrices

In a normalized (or restricted) matrix, the coefficients of each row add to a single fixed value. For example, Equations (I1), (I2), and (I3) show a 3×3 matrix $\boldsymbol{M}$ and the relationships between the matrix

Digital Color Management: Encoding Solutions, Second Edition Edward J. Giorgianni and Thomas E. Madden
 ISBN: 978-0-470-51244-9

Table I.1 Sample pairs of input/output values for the example one-dimensional lookup table illustrated in Figure I.1

Input Value	Output Value
0	0
11	11
22	22
46	46
72	72
91	91
107	107
133	134
155	156
174	175
192	192
206	204
221	213
235	222
247	228
255	232
270	238
292	245
311	250
330	253
346	255

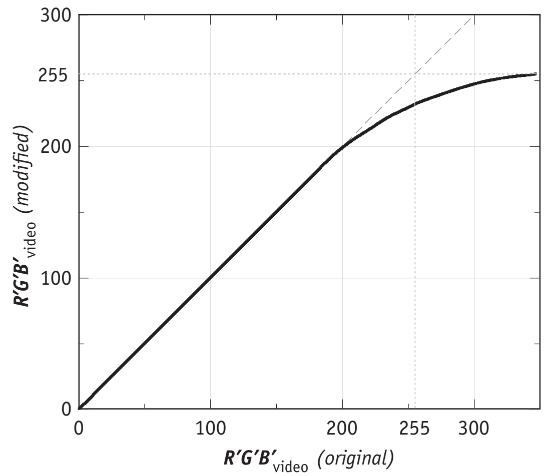


Figure I.1 A graphical representation of an example one-dimensional lookup table. A table such as this might be used to prevent highlight clipping when extended-range input values are transformed to output 8-bit digital code values.

coefficients when the matrix is normalized such that each row sums to the fixed value k.

If

$$M = \begin{bmatrix} m_{11} & m_{12} & m_{13} \\ m_{21} & m_{22} & m_{23} \\ m_{31} & m_{32} & m_{33} \end{bmatrix} \tag{I.1}$$

then

$$\begin{aligned} m_{11} + m_{12} + m_{13} &= k \\ m_{21} + m_{22} + m_{23} &= k \\ m_{31} + m_{32} + m_{33} &= k \end{aligned} \tag{I.2}$$

Accordingly, there are only six independent variables—the off-diagonal coefficients—in a restricted matrix. The three diagonal coefficients are used to adjust the sums of their respective rows, thus they are dependent variables. This relationship is shown by rearranging Equation (I2):

$$\begin{aligned} m_{11} &= k - m_{12} - m_{13} \\ m_{22} &= k - m_{21} - m_{31} \\ m_{33} &= k - m_{31} - m_{32} \end{aligned} \tag{I.3}$$

In imaging applications, this restriction is particularly valuable where matrices are used for color correction and for color-space transformations. When used with a normalized color space (i.e., a space in which neutrals are represented by equal values in all three color channels), a normalized 3 × 3 matrix will have no effect on neutral colors. For example, if a particular neutral has *RGB* code values all equal to 128, passing these values through a matrix that is row-sum normalized to 1.00 will leave all three values unchanged at 128. We highly recommend the use of normalized color spaces and normalized matrices exactly for that reason.

The use of normalized color spaces and 1-D LUTs together with normalized matrices creates a simple and robust system in which neutral signal processing and color signal processing are separable. Normalized matrices can be removed entirely from such systems without affecting the grayscale. This makes it simple and convenient to troubleshoot 1-D problems without the complications created by color interactions.

When applied to color-image data expressed in terms of normalized linear color spaces, normalized matrices can be used to convert from one set of *RGB* primaries to another (see Appendix H). Similarly, conversions can be made to spaces in which luminance and chrominance are separate (i.e., a YCC or YUV space, which is really just a special type of RGB space). Conversion to a luminance/chrominance space is often done prior to implementing image compression, where the degree of spatial subsampling applied to the chrominance channels generally is greater than that applied to the luminance channel.

A normalized matrix operating in linear space will primarily affect higher-chroma colors. That is because, in a linear space, the ratios between color-primary values are larger for such colors. In a nonlinear space, such as a logarithmic space, a normalized matrix will tend to affect most non-neutral colors somewhat independently of their chroma level. In a color-imaging system that includes matrices operating in both spaces, this distinction makes it possible to adjust colors somewhat differently as a function of their chroma level. That capability can be useful to adjust, for example, the hue of skin tones without affecting the hue of higher-chroma reds.

Unnormalized 3 × 3 matrices

In an unnormalized matrix, the coefficient row sums are not restricted to a single fixed value. This means that the matrix can produce different gains (i.e., scale factors) in each row in addition to the interactions produced by the matrix coefficients. An unnormalized matrix combines a 3 × 3 diagonal matrix, which produces the gains, with a normalized 3 × 3, as

$$\begin{aligned}\boldsymbol{U} &= \operatorname{diag}\left(\left[\boldsymbol{g}\right]\right)[\boldsymbol{M}] \\ &= \begin{bmatrix} g_1 & 0 & 0 \\ 0 & g_2 & 0 \\ 0 & 0 & g_3 \end{bmatrix} \begin{bmatrix} m_{11} & m_{12} & m_{13} \\ m_{21} & m_{22} & m_{23} \\ m_{31} & m_{32} & m_{33} \end{bmatrix}\end{aligned} \tag{I.4}$$

where $\boldsymbol{U}$ is the unnormalized 3 × 3 matrix; g_1, g_2, and g_3 are the individual gains for each of the three color channels; and $\boldsymbol{M}$ is the normalized 3 × 3 matrix.

In linear exposure space, changing the gains equally in each color channel is equivalent to changing the overall exposure level. Changing the gains unequally in the channels is equivalent to changing the color balance (or white point). An unrestricted matrix can be used, then, to transform between spaces where the white points are different. One example of this usage would be for what is commonly called a *phosphor matrix*. Such matrices are used to transform *RGB* intensity values to corresponding CRT monitor CIE *XYZ* values. Phosphor matrices are derived based on the chromaticity of each of the three phosphors and the chromaticity of the monitor white. The same matrix derivation procedure can be used for other additive display devices as well when the chromaticities of their primaries and white points are known (see Appendix H).

Another likely use of an unnormalized matrix in imaging systems would be for scanner calibration. Transforming from scanner *RGB* density values to ISO Status M values will almost certainly require individual red, green, and blue gain adjustments in addition to any crosstalk adjustments. Of course this could be accomplished instead using 1-D LUTs and restricted matrices, but that would be necessary only if it is determined that linearity differences, and not just simple gain differences, are involved.

3 × 4 matrices

A 3 × 4 "matrix" is not a true matrix; it is a set of equations consisting of a 3 × 3 linear matrix (restricted or unrestricted) with constants added to each row. Technically, then, each row is an *affine transform*, i.e., a linear transform plus a shift. Thus the 3 × 4 relationship commonly notated in matrix form as

$$\begin{bmatrix} R_2 \\ G_2 \\ B_2 \end{bmatrix} = \begin{bmatrix} m_{11} & m_{12} & m_{13} & c_1 \\ m_{21} & m_{22} & m_{23} & c_2 \\ m_{31} & m_{32} & m_{33} & c_3 \end{bmatrix} \begin{bmatrix} R_1 \\ G_1 \\ B_1 \end{bmatrix} \tag{I.5}$$

is more correctly notated as

$$\begin{bmatrix} R_2 \\ G_2 \\ B_2 \end{bmatrix} = \begin{bmatrix} m_{11} & m_{12} & m_{13} \\ m_{21} & m_{22} & m_{23} \\ m_{31} & m_{32} & m_{33} \end{bmatrix} \begin{bmatrix} R_1 \\ G_1 \\ B_1 \end{bmatrix} + \begin{bmatrix} c_1 \\ c_2 \\ c_3 \end{bmatrix} \tag{I.6}$$

and expands to

$$\begin{aligned} R_2 &= m_{11}R_1 + m_{12}G_1 + m_{13}B_1 + c_1 \\ G_2 &= m_{21}R_1 + m_{22}G_1 + m_{23}B_1 + c_2 \\ B_2 &= m_{31}R_1 + m_{32}G_1 + m_{33}B_1 + c_3 \end{aligned} \tag{I.7}$$

In linear exposure space, a constant of positive value would correspond to an addition of exposure or light, which is equivalent to adding flare light. Using a negative constant value, then, provides a form of flare compensation. This often is a convenient way to compensate for varying amounts of flare, without having to alter any system 1-D or 3-D tables.

In log exposure space, changing the constants equally is equivalent to a change in overall exposure (and equivalent to a gain in linear exposure space). Changing the constants unequally shifts the color balance and is similar to using a colored filter over a camera lens or changing a white-point setting in a digital camera.

Another appropriate application for a 3 × 4 matrix is in the calibration process of negative-film scanners. Transforming from scanner *RGB* density values to ISO Status M values or to printing-density values is likely to involve differences in minimum density (D_{min}) values. Use of a 3 × 4 matrix allows any such differences to be accounted for by the constants. Similarly, the constants can be used to provide densitometric measurements in terms of D_{min}-subtracted values. The use of such values compensates for unwanted film D_{min} variations resulting from chemical process variations and other factors.

When performing a regression that should result only in a 3 × 3 matrix, it is good practice to make an initial run using 3 × 4 equations. If everything is as it should be, the determined constants will be very nearly zero. If so, the regression for the 3 × 3 matrix then can be run with confidence. This is a useful way to verify that there are no unexpected offsets "hidden" in the data.

Polynomial equations

Polynomial equations can be used in color-imaging applications where simple matrices are insufficient to create or compensate for complex color interactions. Such equations might be required, for example, in a transform used to convert film integral density values (e.g., Status M, printing-density, or scanner density values) to channel-independent density values, as required when determining film *RGB* exposure values from film density values.

In addition to an included 3 × 3 matrix, a set of polynomial equations also would have cross-product terms (e.g., red × red, red × green, red × blue), constants, and possibly terms of higher than second order. (Our "personal best" was a set of 3 × 36 equations used in modeling a photographic slide film!) A more typical (and perhaps more reasonable) set of equations, limited to second-order interactions, is shown in the following two equations:

$$\begin{bmatrix} R_2 \\ G_2 \\ B_2 \end{bmatrix} = \begin{bmatrix} m_{11} & m_{12} & m_{13} & m_{14} & m_{15} & m_{16} & m_{17} & m_{18} & m_{19} \\ m_{21} & m_{22} & m_{23} & m_{24} & m_{25} & m_{26} & m_{27} & m_{28} & m_{29} \\ m_{31} & m_{32} & m_{33} & m_{34} & m_{35} & m_{36} & m_{37} & m_{38} & m_{39} \end{bmatrix} \times \begin{bmatrix} R_1 \\ G_1 \\ B_1 \\ R_1^2 \\ G_1^2 \\ B_1^2 \\ R_1 G_1 \\ R_1 B_1 \\ G_1 B_1 \end{bmatrix} + \begin{bmatrix} c_1 \\ c_2 \\ c_3 \end{bmatrix} \tag{I.8}$$

which expands to

$$\begin{aligned} R_2 &= m_{11}R_1 + m_{12}G_1 + m_{13}B_1 + m_{14}R_1^2 \\ &\quad + m_{15}G_1^2 + m_{16}B_1^2 + m_{17}R_1G_1 + m_{18}R_1B_1 \\ &\quad + m_{19}G_1B_1 + c_1 \\ G_2 &= m_{21}R_1 + m_{22}G_1 + m_{23}B_1 + m_{24}R_1^2 \\ &\quad + m_{25}G_1^2 + m_{26}B_1^2 + m_{27}R_1G_1 + m_{28}R_1B_1 \\ &\quad + m_{29}G_1B_1 + c_2 \\ B_2 &= m_{31}R_1 + m_{32}G_1 + m_{33}B_1 + m_{34}R_1^2 \\ &\quad + m_{35}G_1^2 + m_{36}B_1^2 + m_{37}R_1G_1 + m_{38}R_1B_1 \\ &\quad + m_{39}G_1B_1 + c_1 \end{aligned} \tag{I.9}$$

In practice, the values for the various coefficients of such polynomials often are determined by regression techniques. For example, an output device might be characterized by a regression involving its output values and its input digital code values.

In applications where a set of polynomials is to be used to model or to characterize an actual device or medium, it can be advantageous to eliminate some polynomial terms that have no apparent physical significance. For example, it might be quite reasonable to expect the value of an output defined by R_2

in Equations (I9) to be influenced by interactions involving inputs R_1 and G_1 as well as interactions involving inputs R_1 with B_1. However, it would seem far less likely that there would be some real mechanism by which an output would be affected by interactions involving only the other two colors (G^2, B^2, and G_1B_1 in row R_2). Therefore, when the objective is to generate a model that can be related to physical mechanisms, it is reasonable to set the values of such coefficients to zero, as shown below, and not allow them to change during the regression.:

$$\begin{aligned}
R_2 = {} & m_{11}R_1 + m_{12}G_1 + m_{13}B_1 + m_{14}R_1^2 \\
& + 0.0G_1^2 + 0.0B_1^2 + m_{17}R_1G_1 + m_{18}R_1B_1 \\
& + 0.0G_1B_1 + c_1 \\
G_2 = {} & m_{21}R_1 + m_{22}G_1 + m_{23}B_1 + 0.0R_1^2 \\
& + m_{25}G_1^2 + 0.0B_1^2 + m_{27}R_1G_1 + 0.0R_1B_1 \\
& + m_{29}G_1B_1 + c_2 \\
B_2 = {} & m_{31}R_1 + m_{32}G_1 + m_{33}B_1 + 0.0R_1^2 \\
& + 0.0G_1^2 + m_{36}B_1^2 + 0.0R_1G_1 + m_{38}R_1B_1 \\
& + m_{39}G_1B_1 + c_1
\end{aligned} \tag{I.10}$$

Some caution always should be exercised when using polynomials. In fitting measured data, the use of increasingly complex equations generally will yield better statistical fits. However, when there is noise in the data (which is always the case for densitometry and virtually all other forms of measurement), a very complex set of equations may simply be providing a better fit to that noise. A set of less complex equations will tend to smooth over measurement noise, and therefore may provide a more appropriate and useful transform.

3-D LUTs

A 3-D LUT is a cubic lookup table relating triads of input values to corresponding triads of output values. Such 3-D LUTs often do not contain data for every possible triad of input values. In these cases, various forms of interpolation are used to determine appropriate output values for input triads that fall between the actual data points (nodes) of the table.

Commonly, 3-D LUTs are used in image signal processing because they provide fast execution of complex transformations. Depending on their size, 3-D LUTs also can provide an almost unlimited number of degrees of freedom in relating input and output values. That is both their strength and their weakness. It is an obvious strength in that it allows very complex relationships to be characterized. However, that strength also provides opportunities for inappropriate usage.

As with complex polynomials, 3-D LUTs can be "overzealous" in fitting every last wrinkle of noise in a set of measured data. Therefore, it is highly recommended to first fit raw data using a set of polynomials or other equations of no more complexity than necessary for a satisfactory fit. The resulting equations then can be used to compute the data required to construct the 3-D LUT. This procedure not only generates a smoother set of data, but also provides a method of generating values that exactly correspond to the nodes of the table.

Another danger in using such a powerful tool is that it can encourage less thinking and demand less understanding of the actual mechanisms occurring in the imaging process. As a consequence, the availability of 3-D tables may lead to the use of unnecessary transformations. It is not uncommon, for example, to find color-imaging systems where a 3-D LUT is used to transform *RGB* values to CIELAB values that are then transformed by another 3-D LUT to some other set of *RGB* values. The net result, of course, is simply an RGB-to-RGB transformation, which in all likelihood could have been accomplished using far less complex—and probably more accurate—signal processing.

A similar concern is that 3-D LUTs often are used indiscriminately to fit input and output values with no regard as to what the values actually mean. Perhaps the best example of this involves the derivation of transforms used for converting *RGB* exposure values from noncolorimetric films and digital cameras to CIE *XYZ* values. It can take some experience and thought to come to the realization that an appropriately derived 3 × 3 matrix performs this theoretically "impossible" transformation about as well as it can be done. The extra degrees of freedom provided by 3-D LUTs are of little or no value—and can do great harm—in applications such as this where the relationships between the input and output values are neither systematic nor totally predictable. The most appropriate use of such tables is in applications where interactions that are complex, but also systematic, are to be characterized.

Glossary

***a** *b** diagram** A plot of a^* and b^* values of the 1976 CIE $L^*a^*b^*$ (CIELAB) color space.

absorption The transformation of radiant energy to a different form of energy by interaction with matter; retention of light without reflection or transmission.

achromatic Perceived as having no hue; white, gray, or black.

adaptation The process by which the visual mechanism adjusts to the conditions under which the eyes are exposed to radiant energy.

adaptive white A color stimulus that an observer, adapted to a set of viewing conditions, would judge to be perfectly achromatic and to have a luminance factor of unity.

additive color Color formed by the mixture of light from a set of primary light sources, generally red, green, and blue.

advanced colorimetry Colorimetric measurement and numerical methods that include colorimetric adjustments for certain physical and perceptual factors determined according to perceptual experiments and/or models of the human visual system.

affine transformation A transformation of coordinates that is equivalent to a linear transformation followed by a translation.

AgX Silver halide; a light-sensitive crystalline compound used in conventional photographic materials.

average surround An area, surrounding an image being viewed, that has a luminance factor of about 0.20 and chromaticity equal to that of the observer adaptive white; also called a normal surround.

Bayer pattern A particular arrangement of RGB color filters on a grid of photosensors. The pattern is composed of 50 % green, 25 % red, and 25 % blue filters.

bit Contraction of binary digit; the smallest unit of information that a computer can store and process.

block dye A theoretical dye having equal absorption of light at each wavelength within a given continuous range of wavelengths and no absorption at all other wavelengths of interest.

brightness An attribute of a visual sensation according to which an area appears to exhibit more or less light.

brightness adaptation, general The process by which the visual mechanism adjusts in response to the overall luminance level of the radiant energy to which the eyes are exposed.

brightness adaptation, lateral A perceptual phenomenon wherein a stimulus appears more or less bright depending on the relative brightness of adjacent stimuli.

calibration Procedure of correcting for any deviation from a standard.

camera flare Stray light within an image-capture device and/or image sensor that is recorded by that sensor. The amount of camera flare usually is

Digital Color Management: Encoding Solutions, Second Edition Edward J. Giorgianni and Thomas E. Madden
 ISBN: 978-0-470-51244-9

expressed in terms of the amount of exposure it produces relative to the exposure recorded from a white in the scene.

CCD Abbreviation for *charge-coupled device*; a solid-state sensor often used in digital still cameras and scanners to convert light into an electrical signal.

CCIR (Comité Consultatif Internationale des Radiocommunications) Abbreviation for the *International Radio Consultive Committee*, an international television standardization organization, now ITU-R.

CCIR Recommendation 601 A document containing recommended specifications for digital component video, now referred to as Recommendation ITU-R BT.601, or more informally as Rec. 601.

CCIR Recommendation 709 A document containing recommended specifications for high-definition television signals, now referred to as Recommendation ITU-R BT.709, or more informally as Rec. 709.

CD-ROM Abbreviation for *compact disc read-only memory*; a compact disc used for storing digital data for computer applications.

CES Abbreviation for *color encoding specification.*

CGI Abbreviation for *computer-generated imagery*.

channel independent An imaging channel that produces a signal that is detectable and separate from signals produced by other imaging channels.

characterization Procedure of defining the color characteristics for a representative operating model of an input or output device.

charge-coupled device (CCD) A solid-state sensor often used in digital still cameras and scanners to convert light into an electrical signal.

chroma 1. The colorfulness of an area judged in proportion to the brightness of a similarly illuminated area that appears to be white; degree of departure of a color from a gray of the same lightness.

2. A color component of a color video signal.

chroma subsampling A technique for compressing image information, generally for storage or transmission, in which luma (achromatic) information is retained at full spatial resolution while chroma (nonachromatic) information is reduced.

chromatic adaptation The process by which the visual mechanism adjusts in response to the average chromaticity of the radiant energy to which the eyes are exposed.

chromaticity The property of a color stimulus defined by its chromaticity coordinates, such as its CIE x, y, z values.

chromaticity coordinates The ratio of each of a set of tristimulus values to their sum.

chromaticity diagram A plane diagram in which points specified by chromaticity coordinates represent the chromaticities of color stimuli.

chrominance The properties of a color other than its luminance.

CIE (Commission Internationale de l'Éclairage) The *International Commission on Illumination*; the body responsible for international recommendations for photometry and colorimetry.

CIE colorimetry Measurement of color stimuli according to the spectral responsivities of a CIE Standard Observer.

CIE 1931 Standard Colorimetric Observer An ideal colorimetric observer with color-matching functions corresponding to a field of view subtending a 2° angle on the retina.

CIE tristimulus values The values X, Y, and Z, determined according to the color-matching properties of the CIE 1931 Standard Colorimetric Observer.

CIELAB color space A color space, defined in terms of L^*, a^*, and b^* coordinates, in which equal distances in the space represent approximately equal color differences.

CIELUV color space A color space, defined in terms of L^*, u^*, and v^* coordinates, in which equal distances in the space represent approximately equal color differences.

CIEXYZ color space A color space defined in terms of tristimulus values X, Y, and Z, which are determined according to the color-matching properties of the CIE Standard Colorimetric Observer.

cinematographer An expert in the art of lighting and capturing images using electronic cameras or photographic films.

CIS Abbreviation for *color interchange specification.*

clipping Condition where variation of an input signal produces no further variation of an output signal.

CMY/CMYK Abbreviations for *cyan* (C), *magenta* (M), *yellow* (Y), and *black* (K) dyes or inks used in subtractive color imaging.

code value A digital value produced by, or being provided to, an imaging device.

color encoding The numerical specification of color information.

color-encoding data metric The numerical units in which encoded color data are expressed.

color-encoding method Measurement method and signal processing transformation that determine the meaning of encoded color values.

color encoding specification (CES) A fully specified color-encoding scheme, defined by a color-encoding method and a color-encoding data metric, used for encoding color on an individual system. A complete CES also may include specifications for other factors, such as data compression method and data file format.

colorfulness Attribute of a visual sensation according to which an area appears to exhibit more or less of its hue.

color gamut The limits of the array of colors that can be captured by an image-capturing device, represented by a color-encoding data metric, or physically realized by an output device or medium.

colorant A dye, pigment, ink, or other agent used to impart a color to a material.

colorimeter Instrument that measures color stimuli in terms of tristimulus values according to responsivities prescribed for a standard observer.

colorimetric characteristics Referring to characteristics, such as the color-reproduction characteristics of a device, medium, or system, as measured according to standard colorimetric techniques.

colorimetry A branch of color science concerned with the measurement and specification of color stimuli; the science of color measurement.

colorimetry, standard In this book, refers to colorimetric values determined according to current CIE recommended practices.

color interchange specification (CIS) A fully specified color-interchange scheme that includes a complete colorimetric specification, a defined data metric, and a defined set of reference viewing conditions. A complete CIS also may include specifications for other factors, such as data compression method and data file format.

colorist One who adjusts the electronic color signal processing of images during the post-production of a motion picture or in the transfer of photographic images to video.

colorizing A process in which black and white images are altered to include color.

color management The use of appropriate hardware, software, and methodology to control and adjust color in an imaging system.

color-matching functions The tristimulus values of a sequence of visible monochromatic stimuli of equal radiant power.

color primaries, additive Independent light sources of different color (usually red, green, and blue) which may be combined to form various colors.

color primaries, subtractive Colorants, each of which selectively absorbs light of one of the additive primaries. A cyan colorant absorbs red light, a magenta colorant absorbs green light, and a yellow colorant absorbs blue light.

color stimulus Radiant energy such as that produced by an illuminant, by the reflection of light from a reflective object, or by the transmission of light through a transmissive object.

composite profile A single signal processing profile formed by the concatenation of a sequence of two or more individual profiles.

compositing The combining of portions of images or other visual elements from separate sources to form a single image.

compression A process used to reduce the size of data files for storage or transmission.

computer-generated imagery (CGI) The use of computer graphics in motion picture and other applications to create or enhance images.

concatenation Process of combining a sequence of two or more individual signal processing transforms to form a single equivalent transform.

cones Photoreceptors in the retina that initiate the process of color vision.

contrast, objective The degree of dissimilarity of a measured quantity, such as luminance, of two areas, expressed as a number computed by a specified formula.

contrast, subjective The degree of dissimilarity in appearance of two parts of a field of view seen simultaneously or successively.

control voltage, CRT Voltage signal used to modulate beam current, and thus light output, of a CRT.

corresponding colorimetric values Colorimetric values for corresponding stimuli (see below).

corresponding stimuli Pairs of color stimuli that look alike when one is viewed in one set of adaptation conditions, and the other is viewed in a different set.

coupler An organic compound, used in most photographic media, which reacts with an oxidized developing agent to form a dye.

coupler, colored A coupler (see above) that is itself colored.

crosstalk Transfer of information from one color channel to another.

CRT Abbreviation for *cathode-ray tube*.

cyan One of the subtractive primaries; a cyan colorant absorbs red light and reflects or transmits green and blue light.

DAC Abbreviation for *digital-to-analog converter*.

dark surround An area, surrounding an image being viewed, having a luminance much lower than that of the image itself.

dark-surround effect A manifestation of lateral-brightness adaptation; an observer will perceive an image to have lower luminance contrast if that image is viewed in darker-surround conditions.

data metric The numerical units in which a given set of data is expressed.

day-for-night A technique in which scenes are captured during the day but made to appear as if captured at night by using filters, underexposure and other image-modification methods.

daylight A mixture of skylight and direct sunlight.

daylight illuminant An illuminant having the same, or nearly the same, relative spectral power distribution as a phase of daylight.

densitometer A device for directly measuring transmission or reflection optical densities. For meaningful color measurements, the spectral responses of the densitometer must be specified.

densitometry The measurement of optical density.

density, optical The negative logarithm (base 10) of the reflectance factor or transmittance factor.

device-independent color As defined by the authors, refers to techniques for numerically specifying and encoding color information in a way that is not restricted to either the luminance dynamic range or the color gamut achievable by physically realizable devices.

diffuse Referring to light that is scattered, widely spread, not concentrated.

digital color encoding The representation of color information in the form of digital values.

digital quantization Conversion of continuous quantities to discrete digital values; the number of discrete values is determined by the number of bits that are used.

digitize Convert analog signals or other continuous quantities to digital values.

display An image presented to an observer; the process of presenting that image; a device for presenting images.

DLP® Abbreviation for *Digital Light Processor*.

DLP® display A electronic image-projection display device in which light intensities are modulated by a chip containing an array of microscopic reflectors that can be rapidly switched on and off. Images are produced by projecting light reflected by the chip onto a front or back display surface.

DMD Abbreviation for *digital micromirror device*; a semiconductor-based "light switch" array of individually addressable mirror pixels, used in DLP-based displays.

DPX Abbreviation for *Digital Moving-Picture Exchange*, defined in ANSI/SMPTE 268M-1994, SMPTE Standard for File Format for (DPX), v1.0, 18 February 1994.

duplicate A reproduction that is a one-to-one physical copy of an original. The spectral properties of the colorants of a duplicate are identical to those of the original.

dyes Organic colorants used in silver-halide-based photographic media and in other imaging technologies.

dynamic range Extent of minimum and maximum operational characteristics.

encoder and decoder circuits Circuits used in video systems to combine various signals into a composite signal and to subsequently extract the individual signals from the composite.

END Abbreviation for *equivalent neutral density*.

equivalent neutral density (END) Density value used to express the relative amount of subtractive-medium colorants. The END value for a given amount of a colorant is the visual density that would result if appropriate amounts of the other colorants of the medium were added such that a colorimetric neutral is formed, based on a defined viewing illuminant.

exposure The quantity of radiant energy received per unit area; the quantity of radiant energy that is captured by a detector or that forms a detectable signal.

exposure factor Ratio of exposure to that from a perfect diffuser that is illuminated identically.

field That portion of the surface of a specimen that is illuminated by the illuminator or viewed by the receiver.

film terms Input signal processing transforms used on Kodak Photo CD system scanners to convert scanned values to PhotoYCC values.

film writer An output device, used in hybrid color-imaging systems, which produces an image on a photographic film.

flare Stray light; a non-imagewise addition or redistribution of light.

fluorescence Process whereby incident radiant power at one wavelength is absorbed and immediately re-emitted at another (usually longer) wavelength.

gamma correction The use of signal processing in a video camera to complement the characteristics of a video display device such as a CRT.

gamma, CRT 1. Exponent of a power-law equation relating CRT luminance to control signal voltage.
2. The slope of the straight-line portion of a CRT characteristic curve relating log luminance to log voltage.

gamma, photographic The slope of the straight-line portion of a characteristic curve relating optical density to relative log exposure.

gamut, color The limits for a set of colors.

gamut adjustment (or gamut mapping) A method for replacing colorimetric values corresponding to colors that are not physically realizable by a considered output device or medium with substitute values that are attainable by that output. In some methods, values within the attainable gamut also are altered.

gamut boundary Outermost surface of a color space defined by a particular color gamut.

grayscale A progression of achromatic colors from blacks to grays to white.

hardcopy General term referring to solid media such as paper or film base.

HDTV An abbreviation for *high-definition television*, a system having greater spatial resolution than that of previous broadcast television systems.

hue Attribute of a visual sensation according to which an area appears to be similar to one, or to proportions of two, of the perceived colors red, yellow, green, and blue.

hybrid (color-imaging) system A system which incorporates photographic and electronic imaging technologies.

ICC Abbreviation for *International Color Consortium.*

ICES Abbreviation for *input color encoding specification.* One of two specialized color encoding specifications associated with a proposed color-management system for digital cinema. ICES values represent scene-space image colorimetry for a hypothetical reference capture device and a set of reference input-image viewing conditions.

IDT Abbreviation for *input-device transform.*

illuminant A light, which may or may not be physically realizable as a source, defined in terms of its spectral power distribution.

illuminant sensitivity Propensity for colors formed by a set of colorants to change in appearance as the spectral power distribution of the viewing illuminant is changed.

image dyes, image-forming dyes Dyes, usually CMY or CMYK, that make up a displayable image.

independent primaries Sets of light sources in which the chromaticities of each source cannot be matched by any mixture of the remaining sources.

ink A colored liquid or paste used in printing.

input General term referring to imaging media, signals, or data to be put into a color-imaging system.

input compatibility Terminology used by the authors to describe the result of color encoding images such that encoded values completely and unambiguously specify the color of each pixel on a common basis, regardless of the disparity of the sources of the image data.

intensity Flux per unit solid angle; used in this and other texts as a general term to indicate the amount of light.

interlayer effects Chemical reactions that take place among the various layers of a photographic medium. These interactions are used for color signal processing.

International Color Consortium (ICC) An industry group formed in 1993 to promote interoperability among color-imaging systems. The founding members of the consortium were Adobe Systems Inc., Agfa-Gevaert NV, Apple Computer, Inc., Eastman Kodak Company, Microsoft Corporation, Silicon Graphics, Inc., Sun Microsystems, Inc., Taligent Inc., and FOGRA (Honorary). Since its founding, the consortium has continued to expand and now has more than 60 members.

ISO Abbreviation for *International Organization for Standardization.*

isotropic Independent of direction.

ITU Abbreviation for *International Telecommunications Union*; the United Nations regulatory body covering all forms of communication. ITU-R (previously CCIR) deals with radio spectrum management issues and regulation.

ITU Recommendation 601 Recommendation ITU-R BT.601, a document containing recommended specifications for digital component video, formerly CCIR Recommendation 601. Informally referred to as Rec. 601.

ITU Recommendation 709 Recommendation ITU-R BT.709, a document containing recommended encoding specifications for high-definition television signals, formerly CCIR Recommendation 709. Informally referred to as Rec. 709.

latent image A collection of latent-image sites (small clusters of metallic silver within silver halide crystals, formed by exposure of the crystals to light) in a photographic medium. During chemical processing, crystals with latent-image sites are developed to form metallic silver, while those without latent-image sites are not.

lateral-brightness adaptation A perceptual phenomenon wherein a stimulus appears more or less bright depending on the relative brightness of adjacent stimuli. (See dark surround.)

LCD Abbreviation for *liquid crystal display*.

light 1. Electromagnetic radiant energy that is visually detectable by the normal human observer; radiant energy having wavelengths from about 380 nm to about 780 nm.
2. Adjective denoting high lightness.

lightness The brightness of an area judged relative to the brightness of a similarly illuminated area that appears to be white or highly transmitting.

light source A physically realizable emitter of visually detectable electromagnetic radiation, defined in terms of its spectral power distribution.

liquid crystal display (LCD) A thin, flat-panel electronic display device utilizing liquid crystal technology, incorporating monochrome pixels or color subpixels arrayed in front of a light source or reflector. Displayed light intensities are determined by control voltages applied across the liquid crystal layer in each pixel or subpixel.

lookup table (LUT) A computer memory device in which input values act as the address to the memory, which subsequently generates output values according to the data stored at the addressed locations.

luma The achromatic component of a video signal.

luminance A measure, of a luminous surface, that is an approximate correlate to the perception of brightness.

luminance contrast Apparent rate of change from lighter to darker areas of an image. Luminance contrast approximately corresponds to grayscale photographic gamma.

luminance dynamic range Extent of maximum and minimum luminance values, often expressed as a ratio, e.g., 1000:1, or as a logarithmic range, e.g., 3.0 log luminance.

luminance factor Ratio of the luminance of a specimen to that of a perfect diffuser that is illuminated identically.

magenta One of the subtractive primaries; a magenta colorant absorbs green light and reflects or transmits red and blue light.

metadata Descriptive information regarding image file data.

metameric color stimuli Spectrally different color stimuli that have the same tristimulus values.

metameric pair Two spectrally different color stimuli that have the same tristimulus values.

metamerism, degree of Reference to the extent to which matching stimuli are spectrally different. A pair of stimuli that match but have very different spectral characteristics are referred to as being highly metameric.

metamerism, instrument Property of two specimens that measure identically according to the spectral responsivities of an instrument and whose spectral reflectances or transmittances differ in the wavelengths of those responsivities.

metamerism, visual Property of two specimens that match under a specified illuminator and to a specified observer and whose spectral reflectances or transmittances differ in the visible wavelengths.

monitor white Color stimulus produced by a monitor when maximum red, green, and blue code values are applied; measured values for that stimulus.

monochromatic Electromagnetic radiation of one wavelength or of a very small range of wavelengths.

nanometer (nm) Unit of length equal to 10^{-9} meters, commonly used for identifying wavelengths of the electromagnetic spectrum.

negative A photographic medium, usually intended to be printed onto a second negative-working photographic medium, that forms a reversed image, i.e., higher exposure levels result in the formation of greater optical density.

neutral Achromatic, without hue.

normal surround An area, surrounding an image being viewed, that has a luminance factor of about 0.20 and chromaticity equal to that of the observer adaptive white; also called an average surround.

nm Abbreviation for nanometer.

observer metamerism The property of specimens having different spectral characteristics and having the same color when viewed by one observer, but different colors when viewed by a different observer under the same conditions.

OCES Abbreviation for *output color encoding specification.* One of two specialized color encoding specifications associated with a proposed color-management system for digital cinema. OCES values represent rendered image colorimetry for a hypothetical reference display device and a set of reference display viewing conditions.

ODT Abbreviation for *output-device transform.*

OLED Abbreviation for *organic light-emitting diode.*

OLED display An electronic display in which light is produced by special organic materials. Displayed light intensities are determined by control voltages applied across cathode and anode layers, stimulating a middle light-emitting organic layer.

optoelectronic transfer characteristic Characteristic defining the relationship between exposure and output signal voltage for a video camera.

output General term referring to images, signals, or data produced by color-imaging systems.

output rendering The process of converting colorimetric values expressed in terms of a specified scene space to colorimetric values appropriate for display in a specified output-image viewing environment.

PCS Abbreviation for *profile connection space*, a color space used for linking and/or concatenating a series of profiles.

PDP Abbreviation for *plasma display panel.*

perfect white An ideal isotropic diffuser with a spectral reflectance factor or spectral transmittance factor equal to unity at each wavelength of interest.

phosphors Materials, deposited on the screen of a cathode-ray tube, which emit light when irradiated by the electron beam(s) of the tube.

Photo CD player A device, similar to an audio compact disc player, which is used to display images from Photo CD discs on conventional television receivers and monitors.

Photo CD system A hybrid color-imaging system, developed by Eastman Kodak Company, which produces compact discs of images by scanning and digitally encoding images from photographic media.

photographic image-forming dyes The cyan, magenta, and yellow dyes that are formed by the chemical processing of a photographic medium after exposure of that medium to light.

photon A quantum of light or of other electromagnetic radiation.

PhotoYCC color interchange space The data metric of the Kodak Photo CD system, in which color data are encoded in terms of a luma value, Y, and two chroma values, C_1 and C_2.

pigment Finely ground insoluble particles that, when dispersed in a liquid vehicle, give color to paints, printing inks, and other materials by reflecting and absorbing light.

PIW Abbreviation for *Photo CD imaging workstation.* A system consisting of one or more input scanners, computers, monitors, and CD writers used for authoring Photo CD discs.

pixel Contraction of picture element; a single point sample of an image.

plasma display panel (PDP) device A flat-panel electronic display device utilizing plasma technology. Intensities of displayed light are determined by controlling the pulse rate of current flowing through the gas-filled cells of the display panel.

positive A photographic medium, usually intended for direct viewing, in which higher levels of exposure result in the formation of less optical density.

post-production A general term for the various stages of production, such as editing and color adjusting, performed on a motion picture following completion of the principal photography.

power Energy per unit time.

prepress Term used to describe the process, or components of the process, of preparing information for printing after the writing and design concepts stages.

primaries Basic colors used to make other colors by addition or subtraction.

principal subject area The area of a scene that is metered or otherwise used in the determination of camera exposure.

printing density Optical density measured according to a set of effective spectral responsivities defined by the spectral power distribution of a printer light source and the spectral sensitivities of a print medium.

product-specific input transform Input signal processing transforms that can be used to convert input-image values to scene-based color-encoding values. A product-specific transform is based on the characteristics of the particular device or medium being used for input. When product-specific transforms are used, differences among the inputs are minimized in the color encoding.

profile A digital signal processing transform, or collection of transforms, plus additional information concerning the transform(s), device, and data.

profile, input A profile providing the information necessary to convert input-device values to color values expressed in terms of a color-interchange space, color-encoding space or profile connection space.

profile, output A profile providing the information necessary to convert color values expressed in terms of a color-interchange space, color-encoding space, or profile connection space to output-device values.

profile, transform A profile providing the information necessary to modify color values expressed in terms of a color-interchange space, color-encoding space, or profile connection space.

profile connection space (PCS) A color space used for linking and/or concatenating a series of profiles.

psychological, signal processing Modifier used in this book to refer to visual signal processing that includes higher-order mental and cognitive (interpretive) processes.

psychophysical, signal processing Modifier used in this book to refer to visual signal processing that includes both physiological and mental processes.

purple boundary On a CIE chromaticity diagram, the straight line connecting the red and blue ends of the spectrum locus.

quantization Conversion of continuous quantities to discrete digital values; the number of discrete values is determined by the number of bits that are used.

raw (or RAW) file Digital camera image files containing unprocessed data from the camera's image sensor, usually in terms of code values that are proportional to exposure.

Rec. 601 Informal name for Recommendation ITU-R BT.601, formerly known as CCIR Recommendation 601, a document containing recommended specifications for digital component video.

Rec. 709 Informal name for Recommendation ITU-R BT.709, formerly known as CCIR Recommendation 709, a document containing recommended specifications for high-definition television signals.

reference image-capture device A hypothetical reference device defined in terms of spectral responsivities and optoelectronic transfer characteristics and often associated with scene-based color-encoding methods.

reference rendering medium A real or hypothetical imaging medium, the colorimetric properties of which form the basis for color encoding in some types of color-management systems.

reflectance Ratio of the reflected radiant or luminous flux to the incident flux under specified conditions of irradiation.

reflectance factor The amount of radiation reflected by a medium relative to that reflected by a perfect diffuser.

relative colorimetry Colorimetric values expressed relative to those of a reference white. In standard CIE calculations, the reference white is defined to be a perfect white. In "media-relative" colorimetry, the colorimetry of the support of the hardcopy particular medium or the colorimetry of the brightest stimulus produced by an electronic display is defined as the reference white.

rendered image An image having attributes that make it appropriate for display in a specified output-image viewing environment.

rendering, or output rendering The process of converting colorimetric values expressed in terms of a specified scene space to colorimetric values appropriate for display in a specified output-image viewing environment.

retina Layer on the back interior of the eyeball, containing various types of photoreceptive cells that are connected to the brain by means of the optic nerve.

RGB Abbreviation for red, green, and blue.

RRT Abbreviation for *reference rendering transform*. A standardized transform associated with a proposed color-management system for digital cinema. The RRT transforms scene-space image values (ICES) to rendered-image-space values (OCES), which subsequently are used with output-device transforms to produce output-device code values.

saturation The colorfulness of an area judged in proportion to its brightness.

SBA Abbreviation for *scene balance algorithm*, an algorithm that automatically adjusts the overall lightness and color balance of images.

scanner A device for forming image-bearing signals from two-dimensional images.

scene balance algorithm (SBA) An algorithm that automatically adjusts the overall lightness and color balance of images.

scene-based color encoding A method of representing color in terms of colorimetric values and an associated viewing environment having conditions representative of typical outdoor settings.

scene-state image An image defined in terms of its colorimetric values and an associated viewing environment having conditions representative of typical outdoor settings.

sensitivity Property of a detector that makes it responsive to radiant power.

signal processing Chemical, electronic, or digital operations, such as linear and nonlinear amplification, by which original signals are altered and/or combined with other signals.

silver halide A light-sensitive crystalline compound used in conventional chemical photographic media.

simulation The use of one medium or system to imitate the appearance of another.

SMPTE Abbreviation for *Society of Motion Picture and Television Engineers*.

softcopy Jargon for electronic displays such as CRT, LCD, and plasma devices.

source A physically realizable light, the spectral power distribution of which can be experimentally determined.

spatial compression A technique for reducing image information, generally for purposes of storage or transmission.

spectral Adjective indicating that monochromatic concepts are being considered.

spectral power distribution Power, or relative power, of electromagnetic radiation as a function of wavelength.

spectral reflectance The fraction of the incident power reflected as a function of wavelength.

spectral reflection density Reflection density as a function of wavelength; the negative logarithm of spectral reflectance.

spectral responsivity The response of a detection system, such as a scanner or a densitometer, as a function of wavelength. Spectral responsivity is influenced by the spectral power distribution of the illuminant, the spectral filtration effects of various optical components, and the spectral sensitivity of the detector.

spectral sensitivity The response of a detector to monochromatic stimuli of equal radiant power.

spectral transmittance The fraction of the incident power transmitted as a function of wavelength.

spectral transmission density Transmission density as a function of wavelength; the negative logarithm of spectral transmittance.

spectrum locus On a chromaticity diagram, a line connecting the points representing the chromaticities of the spectrum colors.

specular Referring to light that is reflected or transmitted with little or no scattering.

speed Term used in photography to describe sensitivity to light. Higher speed means greater sensitivity to light; lower speed means lesser sensitivity to light.

Standard Colorimetric Observer An ideal observer having visual response described according to a specified set of color-matching functions.

Standard Illuminants Relative spectral power distributions defining illuminants for use in colorimetric computations.

Status A densitometer Densitometer having spectral responsivities corresponding to those specified by the ISO for Status A densitometers. Status A densitometers are used for measurements of photographic and other types of hardcopy media that are meant to be viewed directly by an observer. Status A measurements are *not* equivalent to CIE colorimetric measurements.

Status M densitometer Densitometer having spectral responsivities corresponding to those specified by the ISO for Status M densitometers. Status M densitometers are used for measurements of photographic negative media. Status M measurements are *not* equivalent to printing density measurements.

stimulus, color A spectral power distribution, such as that produced by an illuminant, by the reflection of light from a reflective object, or by the transmission of light through a transmissive object.

subsampling Sampling within samples; a technique employed to compress digital image files.

subtractive color Color formed by the subtraction of light by absorption, such as by cyan, magenta, and yellow (CMY) photographic dyes or by cyan, magenta, yellow, and black (CMYK) printing inks.

surface color Color perceived as belonging to the surface of a specimen, without the specimen appearing to be self-luminous.

surround The area surrounding an image being viewed.

surround effect A manifestation of lateral-brightness adaptation; an observer will perceive an image as having lower or higher luminance contrast depending upon the average luminance of the surround relative to that of the image.

tags In an image file or profile, descriptors of the underlying data.

telecine An imaging system used to scan motion picture films to produce video signals for recording and television broadcast.

test target A collection of color samples used in the evaluation of color-imaging systems, generally made up of spectrally nonselective neutral samples and samples of various colors.

thermal printer An output device that uses heat to transfer dyes to produce images on reflection or transmission media.

transform One or more signal processing operations, used in color-imaging systems incorporating digital signal processing.

transmittance Ratio of the transmitted radiant or luminous flux to the incident flux under specified conditions of irradiation.

transmittance factor The amount of radiation transmitted by a medium relative to that transmitted by a perfect transmitting diffuser.

transparency An image formed on a clear or translucent base by means of a photographic, printing, or other process, which is viewed by transmitting light through the image.

trichromatic Three-color.

trichromatic system A system for specifying color stimuli in terms of tristimulus values based on matching colors by additive mixture of three suitably chosen reference color stimuli.

tristimulus values The amounts of three stimuli, in a given trichromatic system, required to match a particular color stimulus.

tungsten lamp An electric lamp having filaments of tungsten.

tungsten–halogen lamp Lamp in which tungsten filaments operate in an atmosphere of low-pressure iodine (or other halogen) vapor.

Unified Paradigm A color-management concept developed by the authors integrating the encoding and interchange requirements necessary to provide a comprehensive color-managed environment for the imaging industry.

uniform color space Color space in which equal distances approximately represent equal color differences for stimuli having the same luminance.

universal transforms Input signal processing transforms that may be used to convert input-image values to scene-based color-encoding values. A universal transform is based on the characteristics of a reference device or medium of the same basic type as the actual input. When universal transforms are used, differences of the actual input from its associated reference are reflected in the color encoding.

unwanted (spectral) absorption Spectral absorptions of a colorant in portions of the spectrum where ideally there should be 100 % transmission or reflection.

u', v' diagram A uniform chromaticity diagram, introduced by the CIE in 1976, in which u' and v' chromaticity coordinates are used.

viewing conditions Description of the characteristics of a viewing environment that physically alter a color stimulus or that affect an observer's perception of the stimulus.

viewing flare Stray light present in an environment in which an image is viewed. The amount of viewing flare usually is expressed in terms of its amount relative to that of light reflected from, transmitted through, or produced by a white in the image.

visual density Density measured according to a responsivity corresponding to the $\bar{y}(\lambda)$CIE color-matching function.

visual neutral A metameric match to a spectrally nonselective neutral viewed under identical conditions.

von Kries transformation A chromatic adaptation transformation by which changes in chromatic adaptation are represented as adjustments of the sensitivities of the three cone systems.

wavelength In a periodic wave, the distance between two points of corresponding phase in consecutive cycles.

white balance The process of adjusting the RGB signals of a video camera such that equal signals are produced from an illuminated white object.

writer General term for output devices that use photographic films, photographic papers, or other forms of photosensitive hardcopy media.

x, y diagram A chromaticity diagram in which the x and y chromaticity coordinates of the CIE XYZ system are used.

YCC A shorthand notation referring to color signals separated into a luminance (or sometimes luma) channel, Y, and two chrominance (or sometimes chroma) channels.

yellow One of the subtractive primaries; a yellow colorant absorbs blue light and reflects or transmits red and green light.

YUV A shorthand notation referring to color signals separated into a luminance channel, Y, and two chrominance channels, U and V.

zeroing Adjustment of an instrument such that a zero signal value would be obtained when an ideal reference specimen is measured. For example, reflection densitometers generally are adjusted such that a zero-density reading would be obtained if a perfect white diffuser were measured.

Suggested Reading

Color and Vision

American Society for Testing and Materials, *ASTM Standards on Color and Appearance Measurement*, Sixth Edition, American Society for Testing and Materials, West Conshohocken, PA (2000).

Bartleson, C. J. "Measures of brightness and lightness," *Die Farbe* **28**, No. 3/6 (1980).

Bartleson, C. J., and Breneman, E. J. "Brightness perception in complex fields," *J. Opt. Soc. Am.*, ***57*** (1977).

Braun, K. M., and Fairchild, M. D. "Evaluation of five color-appearance transforms across changes in viewing conditions and media," IS&T/SID 3rd Color Imaging Conference, Scottsdale, AZ (1995).

Breneman, E. J. "The effect of level of illuminance and relative surround luminance on the appearance of black-and-white photographs," *Photogr. Sci. Eng.*, **6** (1962).

Breneman, E. J. "Perceived saturation in stimuli viewed in light and dark surrounds," *J. Opt. Soc. Am.*, **67**(5) (1977).

Breneman, E. J. "Corresponding chromaticities for different states of adaptation to complex visual fields," *J. Opt. Soc. Am.*, **A4** (6) (1987).

Daniels, C. M. "Effect of surround on perceived lightness contrast of pictorial images," Masters Thesis, Rochester Institute of Technology, Rochester, NY (1996).

Estevez, O. "On the fundamental data-base of normal and dichromatic color vision," PhD Thesis, University of Amsterdam (1979).

Evans, R. M. *Eye, Film, and Camera in Color Photography*, John Wiley & Sons, Inc., New York (1967).

Evans, R. M. *An Introduction to Color*, John Wiley & Sons, Inc., New York (1970).

Evans, R. M. *The Perception of Color*, John Wiley & Sons, Inc., New York (1974).

Fairchild, M. D. "Considering the surround in device-independent color imaging," *ColorRes. Appl.*, **20** (1995).

Fairchild, M. D. "Refinement of the RLAB color space," *Color Res. Appl.*, **21** (1995).

Fairchild, M. D. *Color Appearance Models*, Second Edition, The Wiley–IS&T Series in Imaging Science and Technology (2005).

Fairchild, M. D., and Berns, R. S. "Image color appearance specification through extension of CIELAB," *Color Res. Appl.*, **18** (1993).

Hunt, R. W. G. "Revised colour-appearance model for related and unrelated colors," *Color Res. Appl.*, **16** (1991).

Hunt, R. W. G. "An improved predictor of colourfulness in a model of colour vision," *Color Res. Appl.*, **19** (1994).

Hurvich, L. M. *Color Vision*, Sinauer Associates, Sunderland, MA (1981).

Jameson, D., and Hurvich, L. M. "Complexities of perceived brightness," *Science*, **133** (1961).

Nassau, K. *Color for Science, Art and Technology*, North-Holland, Amsterdam (1997).

Nayatani, Y., Takahama, K., Sobagaki, H., and Hashimoto, K. "Color-appearance and chromatic-adaptation transform," *Color Res. Appl.*, **15** (1990).

Stevens, J. C., and Stevens, S. S. "Brightness functions: effects of adaptation," *J. Opt. Soc. Am.*, **53** (1963).

Colorimetry

Berns, R. S. *Billmeyer and Saltzman's Principles of Color Technology*, Third Edition, Wiley-Interscience, New York (2000).

CIE Publication 15.2004, *Colorimetry*, Third Edition, CIE, Vienna (2004).

CIE Standard S 014-1/E:2006: Colorimetry Part 1: CIE Standard Colorimetric Observers.

Hunt, R. W. G. *Measuring Color*, Third Edition, Fountain Press (2004).

Judd, D. B., and Wyszecki, G. *Color in Business, Science, and Industry*, John Wiley & Sons, Inc., New York (1975).

Digital Color Management: Encoding Solutions, Second Edition Edward J. Giorgianni and Thomas E. Madden
 ISBN: 978-0-470-51244-9

Ohta, N., and Robertson, A. *Colorimetry: Fundamentals and Applications*, John Wiley & Sons, Inc., Hoboken, NJ (2006).

Wright, W. D. *The Measurement of Colour*, Fourth Edition, Adam Hilger, Bristol (1969).

Wyszecki, G., and Stiles, W. S. *Color Science*, Second Edition, John Wiley & Sons, Inc., New York (2000).

Color Science and Color Reproduction

Bartleson, C. J., and Breneman, E. J. "Brightness reproduction in the photographic process," *Photogr. Sci. Eng.*, **11** (1967).

Bartleson, C. J., and Clapper, F. R. "The importance of viewing conditions in the appraisal of graphic reproductions," *Pr. Tech.* pp. 136–144 (1967).

DeMarsh, L. E., and Giorgianni, E. J. "Color science for imaging systems," *Phys. Today*, September (1989).

Hunt, R. W. G. *The Reproduction of Colour*, Sixth Edition, The Wiley–IS&T Series in Imaging Science and Technology (2004).

Keelan, B. W. *Handbook of Image Quality: Characterization and Prediction*, CRC Press, Boca Raton, FL (2002).

Lee, H.-C. *Introduction to Color Imaging Science*, Cambridge University Press, Cambridge (2005).

Morovič, J. *Color Gamut Mapping*, The Wiley–IS&T Series in Imaging Science and Technology (2008).

Pearson, M. L., and Yule, J. A. C. "Transformations of color mixture functions without negative portions," *J. Color Appearance*, **II**(1) (1973).

Pointer, M. R. "The gamut of real surface colors," *Color Res. Appl.*, **5** (1980).

von Kries, J. A. In W. Nagel (ed.) *Handbuk der Physiologisches Optik*, Vol. **II**, pp. 366–369, Leopold Voss, Hamburg (1911).

Wyszecki, G., and Stiles, W. S. *Color Science*, Second Edition, John Wiley & Sons, Inc., New York (1982).

Yule, J. A. C. *Principles of Color Reproduction*, John Wiley & Sons, Inc., New York (1967).

Digital Color Image Processing

Giorgianni, E. J., Johnson, S. E., Madden, T. E., and O'Such, W. R. *Fully Utilizing Photo CD Images*, Article Nos. 1, 2, and 4, Eastman Kodak Company, Rochester, NY (1993).

Giorgianni, E. J., and Madden, T. E. "Color encoding in the Photo CD System," *Color for Science, Art and Technology*, Elsevier Science, Amsterdam (1997).

"KODAK Photo CD System—-A Planning Guide for Developers," Eastman Kodak Company, Rochester, NY (1991).

Sharma, G. (ed.) *Digital Color Imaging Handbook*, CRC Press, Boca Raton, FL (2002).

Image Compression

Rabbani, M., and Jones, P. W. *Digital Image Compression Techniques*, SPIE Optical Engineering Press, Bellingham, WA (1991).

International Color Consortium (ICC)

Specification ICC. 1:2004-10 (Profile version 4.2.0.0) Image technology colour management—Architecture, profile format, and data structure.

Additional ICC information at: http://color.org.

Photography

Evans, R. M. *Eye, Film, and Camera in Color Photography*, John Wiley & Sons, Inc., New York (1959).

Evans, R. M., Hanson, W. T., and Brewer, W. L. *Principles of Color Photography*, John Wiley & Sons, Inc., New York (1953).

Mees, C. E. K., and James, T. H. *The Theory of the Photographic Process*, Fourth Edition, Macmillan, New York (1977).

Society of Motion Picture and Television Engineers, *Principles of Color Sensitometry*, SMPTE, New York (1963).

Thomas, W. *S.P.S.E. Handbook of Photographic Science and Engineering*, John Wiley & Sons, Inc., New York (1973).

Video

Jack, K. *Video Demystified*, Fifth Edition, Newnes, Oxford (2007).

Poynton, C. A. *Digital Video and HDTV Algorithms and Interfaces*, Morgan Kaufmann, San Francisco (2003).

Recommendation ITU-R BT.601-1, "Encoding parameters of digital television for studios" (formerly CCIR Recommendation 601-1).

Recommendation ITU-R BT.709, "Basic parameter values for the HDTV standard for the studio and for international programme exchange" (formerly CCIR Recommendation 709).

SMPTE Recommended Practice RP 177-1993, Derivation of Basic Television Color Equations.

Sproson, W. N. *Color Science in Television and Display Systems*, Adam Hilger, Bristol (1983).

Index

Digital Color Management: Encoding Solutions, Second Edition Edward J. Giorgianni and Thomas E. Madden