An Introduction to Digital Color Management

Chih-Yuan Yang (楊智淵) and Chiou-Shann Fuh (傅楸善)
Department of Computer Science and Information Engineering
National Taiwan University, Taipei, Taiwan
r91113@csie.ntu.edu.tw and fuh@csie.ntu.edu.tw

Abstract—This paper introduces digital color management, including a simple explanation about light, the meaning of a spectrum, metabolism, colorimetry, RGB space, XYZ space, lateral brightness adaptation. Finally, we illustrate CRT monitor as an example to show the necessary consideration.

I. INTRODUCTION

WHEN devices capture, translate, and display images, many factors affect the color viewed. For example, we may capture a natural scene as original source on a positive, and convert into slide-films. After that, we may show the image on the slide-films by projecting via a slide projector, or scan the slide-films by a slide-film scanner into a computer. If we scan them, we can get the data in digital form and display the image on a monitor, or print them out on white papers to be the sheets of a calendar. In the description above, you can notice that several forms of image are stored and displayed: the original natural scene, the photo on the slide-film, the photo projected, the photo displayed on a computer monitor, and the photo printed on a paper. Therefore, no matter how subtly or obviously, you can suggest that there must be some color changed after the lengthy transfer. Hence, scientists study the properties of human vision and the characteristic of each device for minimizing the color changed and maintaining the color consistency between each device.

II. COLOR STIMULI

The first important issue about color management is the formation of a color. When we see a tree green, an apple red, and the sky blue, what happens to us at that moment? Of course, everyone knows that there is a beam of light fell into the observer's eyes. But the advanced question is that: what is the different property of the different lights making us see different colors? The answer is the spectrum. There are many sensor cells on our retinas stimulated by light waves with special wave length from almost 400 nm to 700 nm. Outside this range, the cells are not affected. Every beam of light has countless photons contributing to a distribution of each wave length, the distribution is called spectrum. The sensor cells translate the photons to current delivered to our brains, and then we see the image with special colors. Therefore, the spectrum (photon distribution) determines the color we see. In fact, the blue color is the response of photons around 400 nm to 500 nm, the green color is around 500 nm to 570 nm, and so on, the order of the remaining color is yellow, orange, and red. Figure 1 is an example of a spectrum.

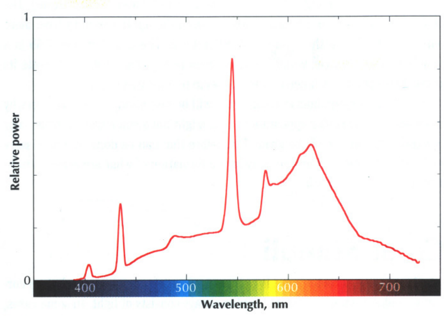


Fig. 1 An example of spectrum.

There are two possible ways of the light we can see. One is that the light emits from the object and into our eyes directly, for example, a bulb, a TV, or a computer monitor. They can emit light themselves. The other way is more popular: objects do not radiate light but reflect the light from others. Therefore, the spectrum we see relies on not only the object, but also the light source. Those two combined properties form the reflected spectrum we receive as shown in Figure 2.

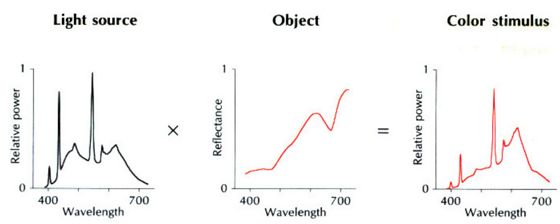


Fig. 2 Reflected spectrum is the product of light source and the object.

III. HUMAN VISION SYSTEM

We can distinguish colors. What is the mechanism in our eyes? In fact, there are two kinds of light sensor cells on our retinas. One kind is like rod, and their ability is to detect the intensity of light. They are very sharp, even in a dim environment, but they do no know what color is because they can only distinguish bright and dark, which is all gray. Oppositely, there is another type of cells on the retina with a different part like a cone from the rod cell, so that they are named cone cells and can distinguish colors, moreover, there are three different types of cone cells with different sensitivities. Long wavelength cone cell is sensitive for red light especially; middle wavelength cone cell is for green, and short wavelength cone cell is for blue. Their relative sensitivities are drawn in Figure 3.

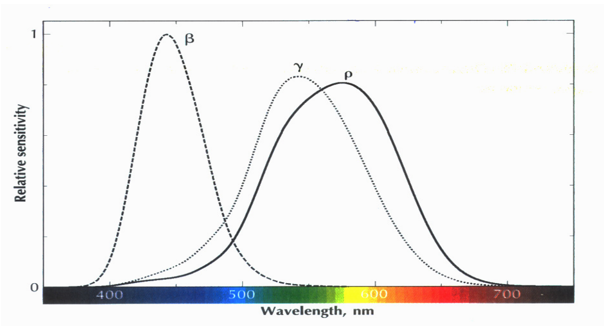


Fig. 3 Relative sensitivity curve of short (β), middle (γ), and long (ρ) wavelength cone cell.

According to Figure 3, we can discover that the sensitivity ranges of middle wavelength cone (γ) and long wavelength cone (ρ) overlap very much, but the difference is broad enough. Hence, when we watch a beautiful flower with vivid color, our long, middle, and short wavelength cones receive the light simultaneously from the flower reflected from the sun. Our long, middle, short wavelength cone cells calculate their stimulus value separately and report to our brains. According to the three values reported simultaneously, our brains judge what the color is. For example, Figure 4 is the spectrum of blue ageratum (a small plant with many purple flowers) and a green fabric.

The response value of each cone cell is the integration of the object's spectrum and the sensitivity curve of each cell. Therefore, we can observe Figure 4(a) and discover that the response of short wavelength cone cell is more intense than middle and long ones. Hence, we see the ageratum blue. On Figure 4(b), middle wavelength cone cell's contribution dominates the total result so that we see the fabric green.

After knowing the property of human vision, we will immediately get an important theory, named metamerism: different spectra may result in the same color we see. Figure 5 shows two different spectra whose energy distribution is quite different, but they provide the same color to us.

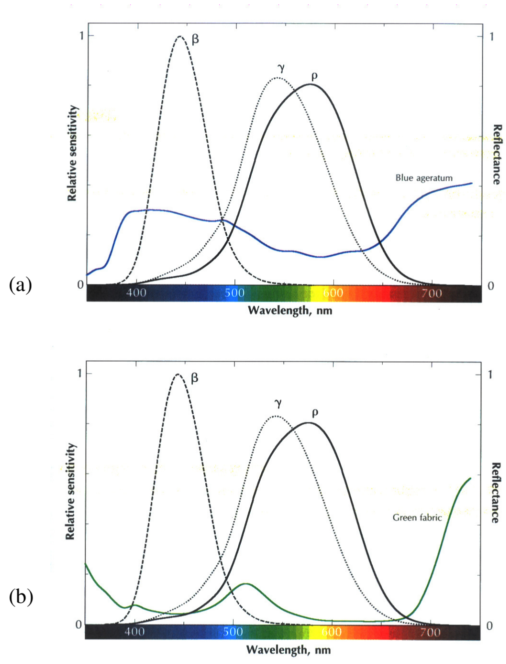


Fig. 4 Spectra of (a) blue ageratum and (b) green fabric.

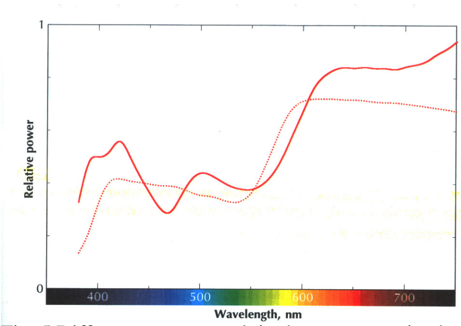


Fig. 5 Different spectra result in the same perceived color.

IV. COLORIMETRY

According to the important property of metamerism, we have a chance to make each color with a precisely numerical form. The experiment started in 1931; hence the result was named CIE 1931. CIE is the abbreviation of Commission Internationale de l'Eclairage (International Commission on Illumination). The experiment system is described briefly in Figure 6.

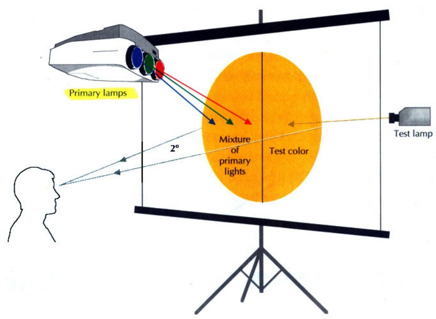


Fig. 6 Condition of colorimetry experiment.

The scientists selected three different kinds of light named primary light with specific wavelength as sources to compare with the test lamp. Their wavelength is 700.0 nm as red light, 546.1 nm as green light, and 435.8 nm as blue light. There is a separating board between projected area of the primary lamp and the projected area of the test lamp to avoid the influence between each side. The observer watched the projected area through a small hole with view angle of 2 degrees. Half was the primary area, and the other half is the test lamp. According to metamerism, we know that we may adjust the intensities of the three primary lamps in order to match the test lamp. When the observer saw the colors of both sides as the same, the scientists recorded the intensity values of the three primary lamps. Therefore, after changing the wavelengths of the test lamp from 400 nm to 700 nm, we can get all the possible metamerism case of the visible light as shown in Figure 7.

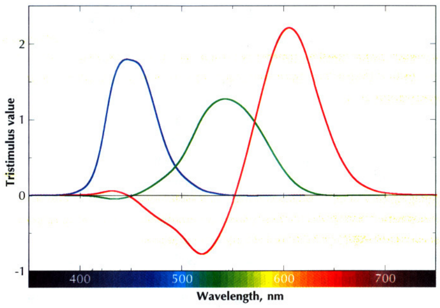


Fig. 7 Metamerism curve of RGB.

But there is one point to be explained. Why are there some negative values in a narrow green segment and a wide red segment? In fact, referring to our preceding description, you can find that there is an assumption in it: every specific wavelength of test lamp can be simulated by the three primary lamps with 700.0 nm, 546.1 nm, and 435.8 nm particularly. Is the presumption true throughout the range of visible light? In fact, it fails at some range. According to Figure 7, we can see that if the test lamp's wavelength is between 450 nm to 550 nm, no matter how we adjust the intensities of the three primary lamps, we can not produce the same color compared with the test lamps. But not to be disappointed too quickly, scientists also discovered that if we stopped using one of the primary lamps in the original side and added it to the other side instead, we can get the metameric balance again as shown in Figure 8.

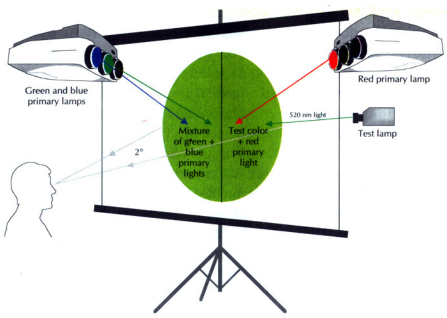


Fig. 8 The meaning of negative values in metamerism curve.

V. RGB COLOR SPACE

Therefore, when we use one of the primary lamps at the same side with the test lamp, we record the value of the switched primary lamp negatively. Up to now, we know that color is the perception of visible light, and the spectrum of visible light can be viewed as the combination of many test lamps with particular wavelength. Furthermore, by metabolism, we can describe every test lamp with three corresponding value. Consequently, we can describe every color in three metameric numerous value on the bases of the primary light source of 700.0 nm, 546.1 nm, and 435.8 nm. Therefore, it forms a color description system, named RGB space. Some people may regard RGB space the three [0, 255] intensity values used in some image process software, for example, Photoshop. Actually, they are different things with the same names.

VI. CIE COLORIMETRY

Though the fundamental RGB color space is usable, it is inconvenient for both general user and color scientists because the negative value of a light is not intuitive. Hence, color scientists had strong motivation to construct a more convenient, intuitive, and useful color space. Therefore, here comes the CIE XYZ color space. As the name of RGB space, the word XYZ means another set of three primary lamps used as colorimetry light source, but light sources X, Y, and Z are all virtual light sources. They do not exist actually; they are created for convenience only. Every set of RGB values can be transferred into XYZ values using the following formula.

$$\begin{aligned} X &= 2.7689\,R + 1.7517\,G + 1.1302\,B \\ Y &= 1.0000\,R + 4.5907\,G + 0.0601\,B \\ Z &= 0.0565\,R + 5.5943\,B \end{aligned} \qquad [1]$$

As a result, the new XYZ values will create some new good properties as follows:

1. For every color, the corresponding XYZ values are all positive.
2. The XYZ values of the light with wavelength from 540 nm to 700 nm fall on a straight line through X and Y; therefore, in the range from 540 nm to 700 nm, it is sufficient to describe the color by X and Y only.
3. Y values can present the brightness of a light.
4. A white light with equal energy distribution from 400 nm to 700 nm will be at the center of the three points, e.g. (X, Y, Z) = (1/3, 1/3, 1/3).

Furthermore, as shown in Figure 9, after projecting all (X, Y, Z) points from pure colors from 400 nm to 700 nm to the unit plane of $X + Y + Z = 1$ by the following formula.

$$x = \frac{X}{X + Y + Z} \qquad [2]$$
$$y = \frac{Y}{X + Y + Z}$$

We can get a horse-hoof-shape locus on the *xy* plane as shown in Figure 10, which will be very useful to discuss the capacity of a display device.

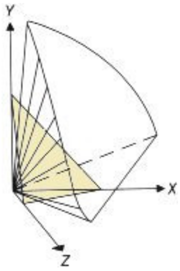


Fig. 9 Projection *XYZ* space to *xy* plane.

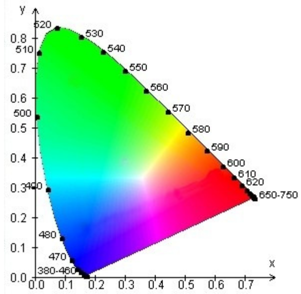


Fig. 10 Gamut, the projected result on *xy* plane.

VII. Adaptation

Up to now, all things seem simple and analytical. If we want to compare two colors, it seems enough to get their XYZ values then we can distinguish them completely. But in fact, this is enough because color is not only a physical effect but also a psychological effect. For example, in Figure 11, there are six gray squares. Each pair in the same vertical position has the same brightness, but you still feel they are different. Especially, you may feel that the left ones are darker than the right ones. The effect is called lateral-brightness adaptation. Beside lateral-brightness, there are general-brightness adaptation and chromatic adaptation, which affect our perception of color whenever.

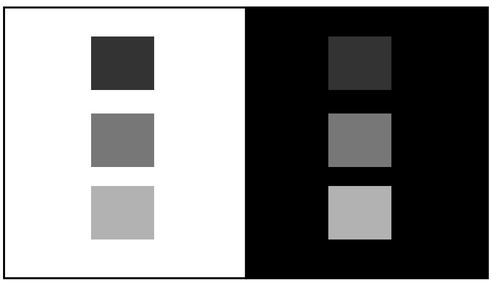


Fig. 11 Example of lateral-brightness adaptation.

Therefore, for achieving the result of color consistency, the target of digital color management, we have to keep the adaptation effect in mind all the time.

VIII. Example: Video Image

First, we have to define what a video image is. For example, video image is like an image shown on a TV monitor, a CRT (Cathode Ray Tube) monitor, or a LCD (Liquid Crystal Display) panel. All devices which can emit light ray and control the intensity are called video image devices, and they work similarly. Therefore, we illustrate CRT monitor, a representative device in the category, in this paper.

At the beginning, it is necessary to know how a CRT operates. Figure 12 illustrates mechanism inside a CRT monitor.

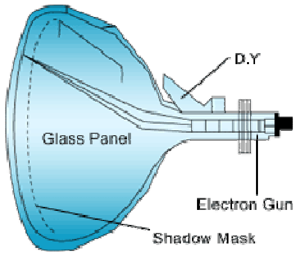


Fig. 12 Mechanism of a CRT monitor.

From the narrow neck bottom to the outspread glass panel, electrons are accelerated by electron gun and the direction was controlled by the deflection yoke. Furthermore, a phosphor layer is painted inside the glass panel. When the electrons collide with the phosphor layer, the phosphor particle's energy level rises and become unstable. When the energy level fall down to normal level immediately, the phosphor particle emits light we see. Figure 13 shows the spectra of those three phosphors.

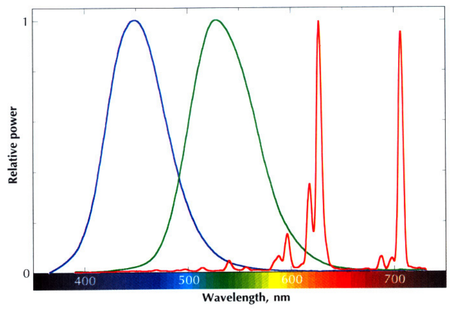


Fig. 13 Spectra of CRT's phosphors.

Besides, let us ask a question. After receiving the RGB code as input, what is the luminance intensity the CRT monitor generates on the glass surface? Many people think it is a linear function, but in fact, it is an exponential function as shown in Figure 14.

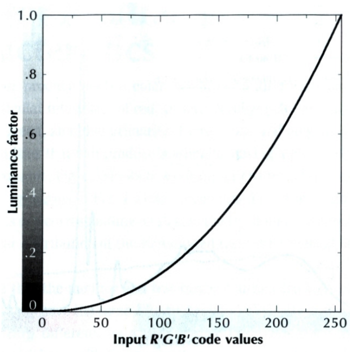


Fig. 14 The gamma curve of input code and output luminance.

The function can be described in Equation 3.

$$Y = k\left(\frac{C_{RGB}}{255}\right)^{r} + Y_0 \qquad [3]$$

The value of Y_0 is not zero because of the characteristic of the electron gun of the CRT. When the input signal, C_{RGB} in Equation 3, is zero, ideally no electrons are accelerated to collide with the phosphor. But in fact, because of the very high temperature of the electron gun, there are always some electrons escaping and hitting on the phosphor.

For every CRT monitor, after knowing the values of normalizing factor k, the exponent r (gamma), and the offset Y_0, we can derive an inverse gamma function to achieve the aim of produce any luminance of the phosphor we want, as shown in Figures 15 and 16.

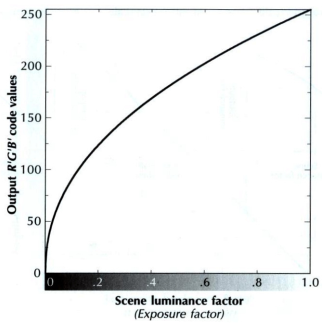


Fig. 15 The inverse gamma function of CRT monitor.

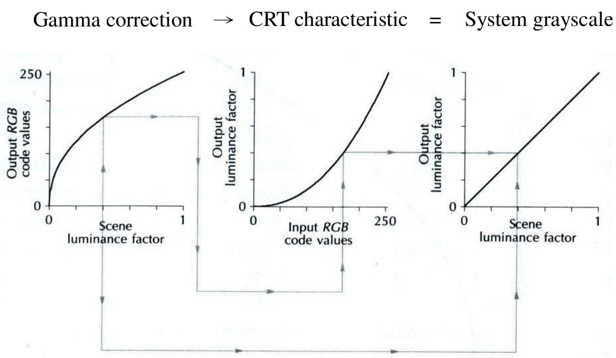


Fig. 16 The compensation of CRT characteristic.

The compensation is called gamma correction and the compensation curve is called gamma curve. For example, if the Y_0 is 0.001 (compared with the CRT's maximal luminance as 1) and the r (gamma) is 2.22, then we can get the inverse function in Equation 4.

$$C_{RGB} = 255\left(\frac{Y - 0.001}{0.999}\right)^{0.45} \qquad [4]$$

Up to now, we can control any exact brightness of each phosphor. Next, we will mix their spectra to show most colors. First, let us see an example. Figure 17 has two curves, the dotted one presents a gray patch of GretagMacbeth ColorChecker under certain light source (D65), and the solid one is produced by CRT monitor.

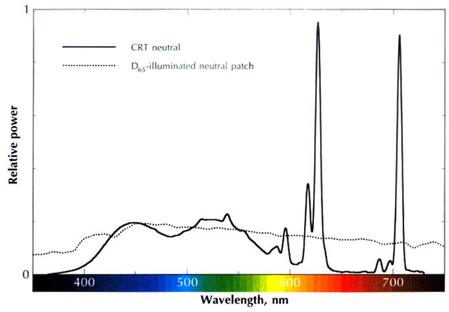


Fig. 17 Identical gray by CRT phosphors and a test patch.

Now, let us recall the purpose of color management: to keep color consistent. This purpose is equivalent to producing identical color on a CRT in this case. Therefore, after knowing the characteristic of a CRT monitor, we can develop a system to produce the color again.

According to the description of Section VI, we know every color has its own XYZ value. Consequently, it makes sense to reproduce a color with the same XYZ values by the phosphors. To achieve it, we have to know the XYZ values of the three phosphors at first. Suppose we know that (X, Y, Z) = (0.4997, 0.2635, 0.0315) when the input signal (R_m, G_m, B_m) = (255, 0, 0), (X, Y, Z) = (0.3163, 0.6548, 0.1390) as (R_m, G_m, B_m) = (0, 255, 0), and (X, Y, Z) = (0.1839, 0.0817, 0.8296) as (R_m, G_m, B_m) = (0, 0, 255). After normalizing (R_m, G_m, B_m) to (I_R, I_G, I_B) with the range from 0 to 1, we can express them in matrix form in Equation 5 because XYZ is a linear color space translated from RGB color space.

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} 0.4997 & 0.3163 & 0.1839 \\ 0.2635 & 0.6548 & 0.0817 \\ 0.0315 & 0.1390 & 0.8296 \end{bmatrix} \begin{bmatrix} I_R \\ I_G \\ I_B \end{bmatrix} \qquad [5]$$

Therefore, for any XYZ values, we can figure out the corresponding phosphor intensities (I_R, I_G, I_B) value from the inverse matrix shown in Equation 6.

$$\begin{bmatrix} I_R \\ I_G \\ I_B \end{bmatrix} = \begin{bmatrix} 2.6542 & -1.1819 & -0.4720 \\ -1.0780 & 2.0398 & 0.0381 \\ 0.0798 & -0.2969 & 1.2169 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} \qquad [6]$$

By combining the gamma correction, CRT grayscale characteristic, and the matrix operation, we can show the total process to generate the color we wish as Figure 18 shows.

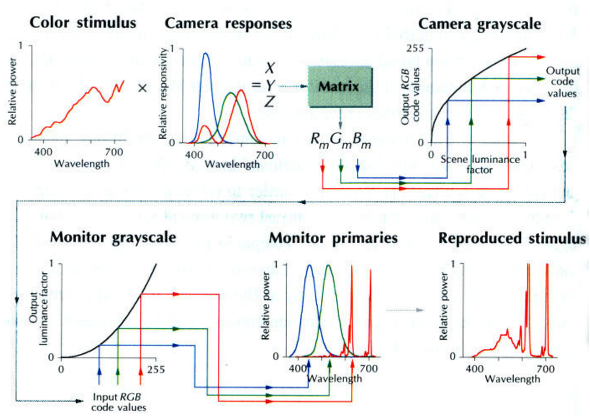


Fig. 18 The process of reproducing the same color.

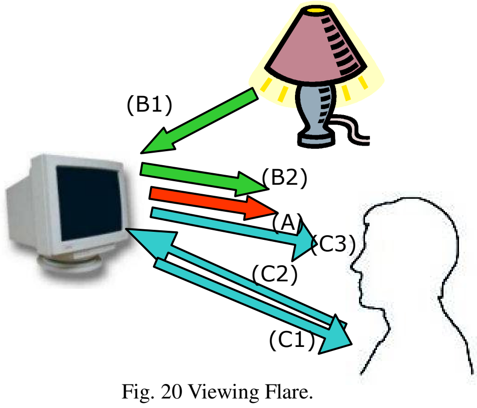


Fig. 20 Viewing Flare.

Now, we have to explain an important fact in color reproduction. Not all colors can be reproduced on a CRT monitor! You can observe that in Equation 6, there are some negative elements in the matrix. Thus, for example, if we want to reproduce a color with (X, Y, Z) = (0.6597, 0.6820, 0.0900), a common vivid yellow, the corresponding (I_R, I_G, I_B) will be (0.9024, 0.6834, -0.0403). Please take care with the last value, which is negative. Recall the introduction of Section IV; you can explain the meaning of a negative number is to illuminate the light of blue phosphor on the compared target. But actually, it is impossible for a CRT to do so. Therefore, you can imagine that the high light reflection of a gold bracelet or of a CD (Compact Disk) can not be represented faithfully. Furthermore, as shown in Figure 19, a CRT monitor can only display colors within the triangle whose vertices are the corresponding *xy* values from Equation 5.

The reason is quite simple. What we can see on the screen of a CRT monitor is the mixture of the three phosphors, and their coordinates can be marked as the vertices in Figure 18. Therefore, the color we can see must be in the triangle.

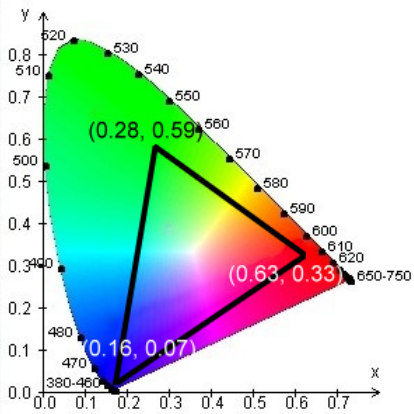


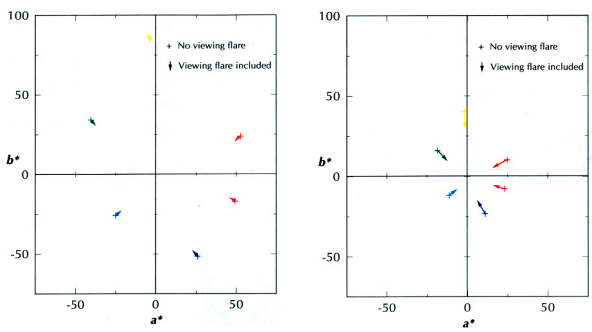
Fig. 19 Color limitation of a CRT monitor.

IX. Viewing Flare

For displaying colors more accurately in real operating environment, we have to consider other factors outside the device such as viewing flare. Figure 20 illustrates viewing flare.

Ideally, if there is no other light interfering, we will only see the light emitted from CRT monitor's glass surface (A) as accurately as we wish. But actually, other light sources exist in our rooms, our offices, and almost everywhere. Therefore, we see from the monitor not only the phosphor light but also others reflected from diverse light sources (B1) (B2). Besides, even if we are in a completely dark room without any light source except the CRT monitor, there still is some viewing flare because the light emitted from the CRT not only shines on our eyes but also on our faces and clothes (C1). Thus, the color we see contains the reflected light from our faces and clothes (C3).

In Figure 21, we show the effect of viewing flare. There are six arrows in each graph. The start-node of the arrow presents the original coordinate in CIELAB color space. The color on the origin is gray, and the distance from the origin means the saturation. The six arrows in parts (a) and (b) are photographed with the same patches, and the only difference is their luminance. Part (b) is lower than part (a). We can see that the effect of viewing flare fades color toward gray more with lower luminance on CRT screen. It makes sense because viewing flare adds a random light on the original color.



(a) (b)

Fig. 21 Influence of viewing flare on saturation.
(a) With brighter illumination. (b) With darker illumination.

To compensate it, we have to measure the intensity of viewing flare at first. As Figure 22 shows, we can post a white patch at the corresponding position on the surface of the monitor. The white patch should obstruct the light from CRT to the observer, thus the value measured by the light meter should be viewing flare at the corresponding position.

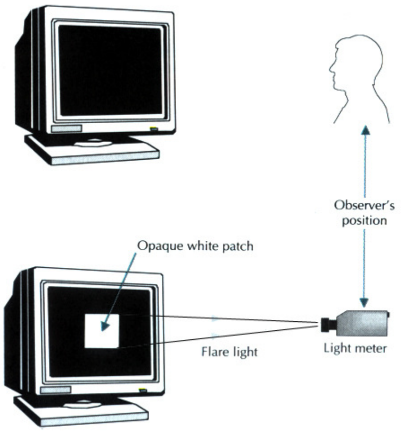


Fig. 22 Measurement of viewing Flare.

After getting the intensity of viewing flare, we can compensate it in the gamma curve by shifting down the value of gamma curve. But you can ask an advanced question: if the intensity is different in different area on the surface of CRT monitor, could we control the uniform gamma curve to solve this problem? The answer is no. We can not solve this problem.

X. Color Management System

Although we can not solve all problems to get the completely accurate reproduced color on a CRT screen, the color is accurate enough for usual use. Therefore, for communicating efficiently with other devices manufactured by different companies, we store the CRT's characteristic into a profile referenced by the color management system as shown in Figure 22. When we change our CRT monitors, the system can adjust to produce appropriate colors by the information from the new device's profile.

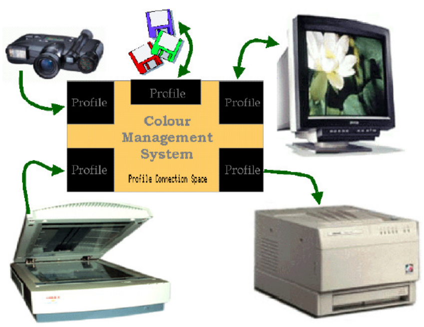


Fig. 22 Color management system

XI. Conclusion

After the illustration in this paper, you may know several things:

1. The meaning of a spectrum.
2. The relation between RGB and XYZ color spaces.
3. The meaning of the horse-hoof-shape gamut on *xy* plane.
4. Colorimetry
5. Adaptation.
6. Gamma curve and its compensation of a CRT monitor.
7. The influence of viewing flare.

Color management attempts to keep color consistent on different devices. Through our knowledge of each device, we try to control their behavior as exactly as possible. Besides display monitors, other devices such as cameras, scanner, and printers also need to be managed. You can imagine that the methods used on different devices are very different fundamentally.

Reference

[1] E. J. Giorgianni and T. E. Madden, *Digital Color Management*, Addison-Wesley, Reading, MA, 1998.
[2] J.Y. Hardeberg, *Acquisition and Reproduction of Color Images*, Dissertation.com, USA, 2001.