

Chapter 6

Color Reproduction

Take a digital camera and click a picture of a scene. This is the *color reproduction* of the original scene. The success of a color reproduction lies in how close the reproduced scene is to the original scene. Note that the term ‘close’ can be measured both quantitatively and qualitatively. There are two types of color reproduction system, additive and subtractive and we will discuss both in details.

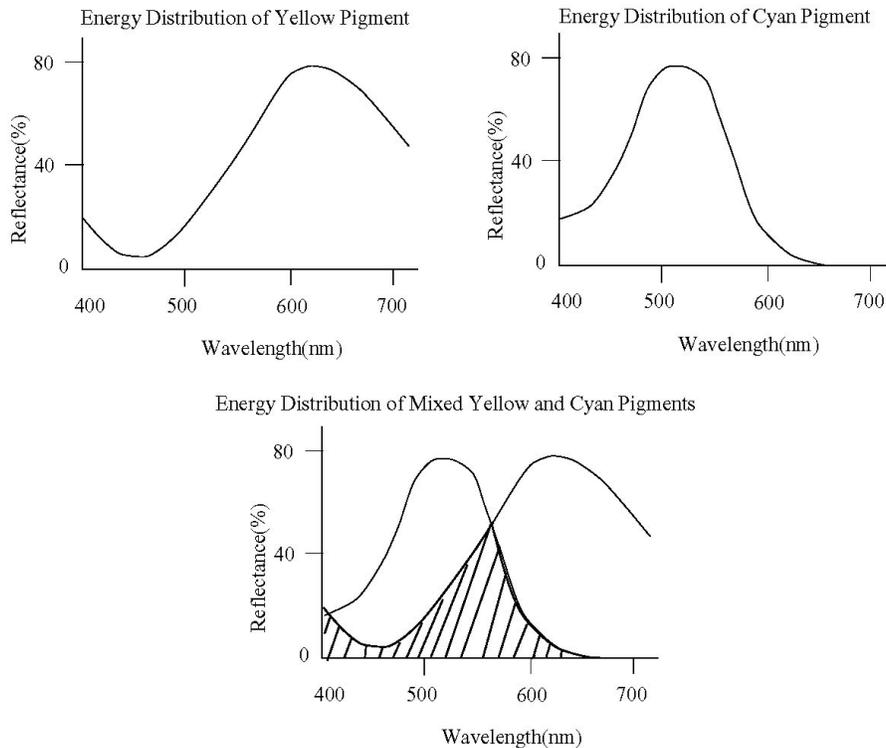


Figure 6.1: Subtractive Color Mixture

In subtractive color mixture, the color of a surface depends on the capacity of the surface to reflect some wavelengths and absorb others. When a surface is painted with a pigment or dye, a new reflectance characteristic is developed based on the capacity of the pigment or dye to reflect and absorb the different wavelengths of light. Consider a surface painted with *yellow* pigment which reflects wavelengths $570 - 580nm$ and another surface painted with *cyan* pigment which reflects $440 - 540nm$. If we mix both the pigments the resulting color will be

green. This is because the yellow pigment absorbs the shorter wavelengths below $500nm$ and some of the middle band wavelength from $500 - 550nm$. The cyan pigment absorbs all of the longer wavelengths $560nm$ and above. The energy distribution of all these are shown in Figure 1.1. Thus the yellow absorbs the wavelengths evoking the sensation of blue while the cyan absorbs the wavelengths evoking the sensation of yellow. Hence, what is left behind after this is a sensation of green. This is called *subtractive color mixtures* since bands of wavelengths are subtracted or cancelled by the combination of *light absorbing materials*. And the resulting color, as you have probably noticed, is given by the *intersection* of the two spectrums. The yellow, cyan and magenta is termed as the color primaries of the subtractive color mixtures, because these are the minimal number of pigments required to produce all other colors. Dyes and inks usually follow subtractive color theory and hence images generated using these mediums are the result of subtractive color reproduction.

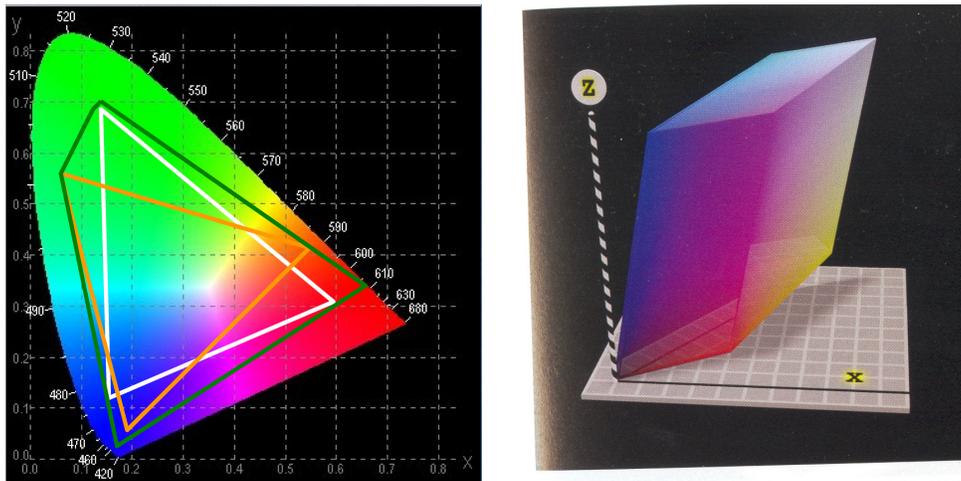


Figure 6.2: Left: Representation of color gamuts on the chromaticity diagrams. Right: The corresponding 3D color gamut.

In additive color mixture systems, colors are mixed in a fashion in which bands of wavelengths are added to each other. This is called *additive mixture of colors*. Thus, the spectrum of the color formed by superposition of multiple colors is given by the *addition* of their respective spectrums. This is similar to how the human eye visualizes color. Devices like camera, projectors follow additive color mixture.

The basic principles of good quality color reproduction are the following.

1. *Correct mapping of critical reference colors* such as sky, foliage and skin tones. This may not mean an exact match but simply that the reproduced color is not grossly wrong. For example, almost any shade of blue will produce a satisfactory sky; even shades of purple would be fine, but green is clearly wrong.
2. *Correct mapping of white and neutral colors* that constitute the gray axis or the neutral axis which runs from black to white. These colors should look neutral, else the image will have an overall *color cast*, or an overall color tint.
3. *Control of the tone reproduction* involves mapping of the overall contrast and brightness. Image reproduction often involve tone compression. The goal is to reproduce, as best as possible, detail at all levels of brightness throughout the image while maintaining a correct overall appearance.
4. *Control of the overall colorfulness* so that the image does not look washed out or gaudy.
5. *Control of sharpness, texture and other visual artifacts* that contribute to image appearance.

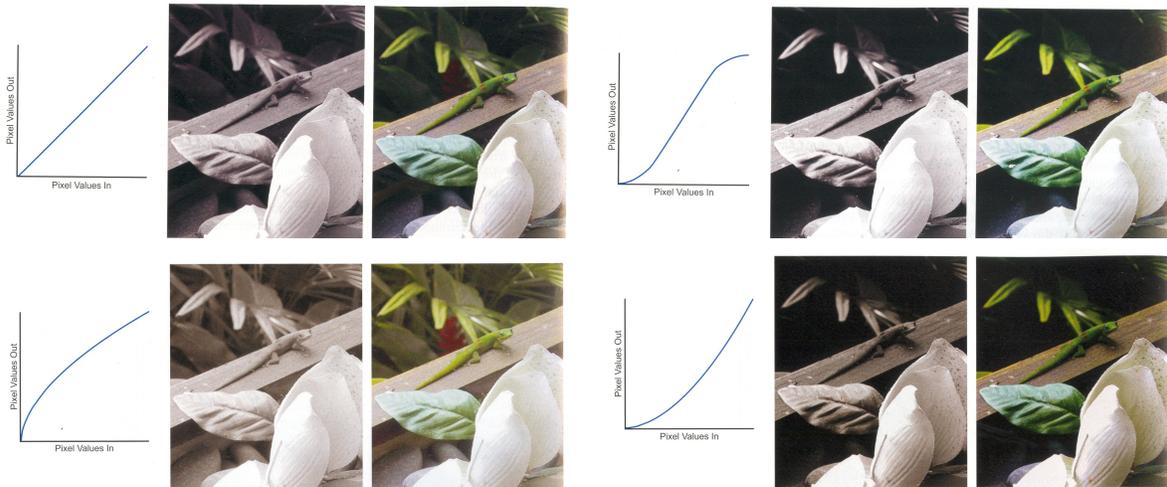


Figure 6.3: Tone Mapping Functions and its effects on color reproduction.

Traditional color reproduction was a linear process. However, in digital color reproduction, this linear model is replaced by a star shaped model. In traditional color reproduction, it was sufficient to capture images in one medium and then tune it to be output to one specific media. This is no longer true. In digital color reproduction, images can be captured in different ways and then tuned to be output in different media. Hence, it is often essential to go through a *device independent* color description like the CIE XYZ color space.

6.1 Additive Devices

Any additive color display is characterized by the following parameters.

6.1.1 Color Gamut

The color gamut of a device describes the entire range of color that can be reproduced by the device. A color device usually have three primaries which can be plotted on the chromaticity diagram as shown by the white triangle in Figure 1.2. This white triangle represents all the hues of different saturation than can be generated by the device. The brighter and more saturated the primaries, the bigger the color gamut. However, note that it is not possible to cover the entire range of colors seen by humans even by choosing three monochromatic primaries (the most saturated colors). Also, note that this representation of the gamut on the chromaticity diagram does not contain any information about the minimum and maximum luminance that can be reproduced by the display. It only signifies the hue and saturation of the primaries, each of which can have different minimum and maximum brightness. So, to complete the description, we need the information about *white point* and the *dynamic range*(more commonly called contrast). The white point gives the chromaticity coordinate of the white and the dynamic range is given by the ratio of the brightest and dimmest gray, i.e. white and black. However, note that dynamic range can provide you the color gamut only upto a scale factor and hence the display's maximum luminance (usually produced at white) is required to get the absolute color gamut. This is commonly called the brightness of the display. All these parameters together define a 3D color gamut in the XYZ space as shown in Figure 1.2.



Figure 6.4: Left: An intensity transfer function maps input values to output intensities. The black dashed curve, starting from black and progressing smoothly to white, is desired. The red curve has brightness set too low such that input values till 0.4 will be displayed as black. The blue curve has brightness set too high (black is not zero). The yellow curve has both brightness and contrast set too high (black is not zero and everything above 0.8 is white). Right: The effects of the different transfer functions. From left to right: (a) Good reproduction (black curve), (b) too dark (red curve), (c) black too high (blue curve) and (d) burned out highlights, overall too light (yellow curve).

6.1.2 Tone Mapping Function

The second parameter that characterizes a device is its *tone mapping* function. This defines the way the input intensity for each primary map to the output intensity. This in turn affects the brightness and contrast which impact the overall appearance and how much detail will be visible. Figure 1.3 shows the effect of different tone mapping functions on a gray and color image. For the color image, the same tone mapping function is applied to all the three channels. Thus, this not only controls the brightness and contrast of the images reproduced but also has an effect on the saturation of the colors achieved. Often tone mapping functions are based on some power function of the form $I_o = I_i^\gamma$. In case of displays, this tone mapping function is called the *channel transfer function*. For optimal image quality, this should be a monotonic smooth function with no flat spots. Most display systems have controls named *brightness* and *contrast* that manipulates these two properties of the display by changing the transfer function. The brightness control usually raises the entire transfer function (by adding an offset to it), whereas the contrast stretches only the maximum brightness. The way these controls change the shape of the transfer function is shown in Figure 1.4. The effect of changing the transfer function on the image reproduction can be dramatic as illustrated in Figure 1.4.

To get the best reproduction, the image should have pixels throughout the entire tone range, to take full advantage of the dynamic range of the medium. This can be visualized by plotting the histogram of the image. If the image does not have a good distribution across the entire tone range, stretching the histogram to map the darkest pixels to black and the brightest to whites is the typical first step to achieve a good tone mapping. This is illustrated in Figure 1.5. The jagged appearance of the histogram is common in this kind of processing.

6.1.3 Color Balance

Color balance describes the relative weights of the color primaries and hence controls the hue and the relative brightness and saturation of colors. Color balance can be adjusted by both changing the tone mapping functions of the different channels differently and also by changing the relative intensities of the primaries using the brightness control. This effect is illustrated in Figure 1.6.

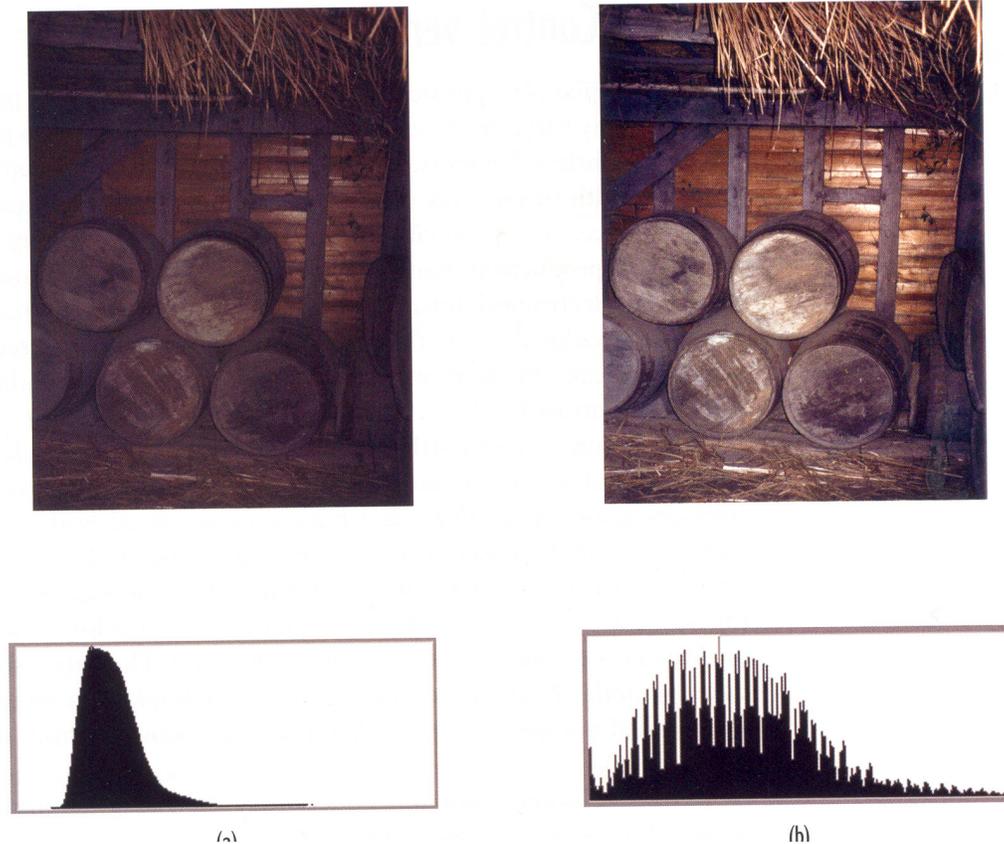


Figure 6.5: Stretching the histogram to achieve better tone mapping.

6.1.4 Intensity Resolution

This ideally means the number of visible intensity steps. However, this is complex function of viewing environment, color and absolute brightness. Instead, intensity resolution is defined by the number of digital steps used to define the intensity of each channel. Thus, for 8 bit display, intensity resolution is 256. The distribution of these intensities depend on the transfer function. Insufficient intensity resolution introduce quantization artifacts in the form of contour lines as illustrated in Figure 1.7. In practice, a perceptually uniform brightness distribution rarely shows contouring with 8 bits per pixel. However, linear encoding on such displays demand 10-12 bits per pixel.

6.1.5 CRT Displays

CRT monitors excite phosphors with rays of electrons from an electron gun. Different types of phosphors are using to emit red, green and blue light. The phosphor colors match the sRGB color gamut, but they age easily becoming progressively less bright. In addition, the blue phosphor often degenerates faster leading to a change in color balance, giving the monitor a yellowish appearance. Note that though the hue and saturation of the primaries remain same in the chromaticity diagram, the color of the display changes just due to deterioration of their brightness.

The CRT's transfer function is a non-linear power function which is defined by the physics of the electron gun exciting the phosphors. In the simplest form, it is approximated as

$$I = V^\gamma$$



Figure 6.6: Effects of changing the color balance.

where I is the measured intensity and V is the input voltage corresponding to channel inputs. If the intensity produced at V_0 (black) is non-zero, then this equation is modified to

$$I = (V + V_0)^\gamma$$

Finally, the entire curve can be scaled by a constant k to give the most general form of the equation

$$I = k(V + V_0)^\gamma$$

The brightness and contrast controls in CRTs therefore change V_0 and k respectively.

6.1.6 Liquid Crystal Displays (LCD)

LCD displays are spatial array of red, green and blue segments, each of which is a colored filter over a cell of liquid crystal material that can be made variably transparent. A backlight shines through the LCD array so that the resulting color is the function of both the backlight and the filters. However, colored filters are significantly different than colored phosphors. The more intensely colored (saturated) the filter is, lesser light will pass making the display dim. The less saturated the filters, brighter is the display but a lot less colorful. Thus, to get a highly bright and colorful display, very bright backlights are needed in addition to saturated color filters. But a trade-off needs to be made due to the huge power consumption. When compared to the gamut of the CRT displays, the blue of the LCD displays are often much less saturated.

The transfer function of LCDs are usually linear. But usually LCDs include electronics by which this can be changed to match the traditional CRT displays. Also, usually LCDs project some light even at input zero due to leaking of the backlight through the front of the display. This is sometimes called flare or black offset. Flare

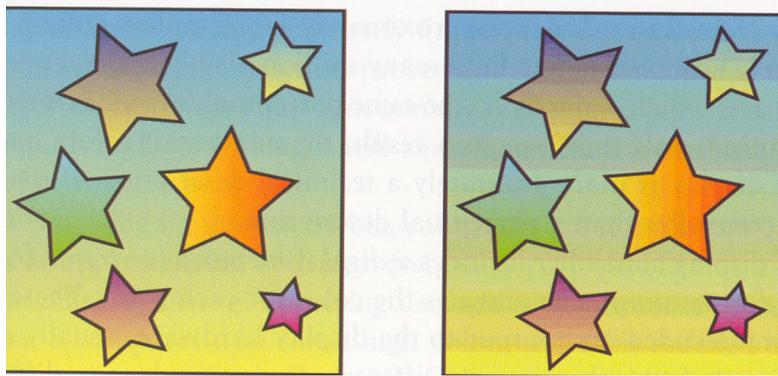


Figure 6.7: Contour lines caused by insufficient intensity resolution. Left: Using 8 bits/pixel. Right: Using 4 bits/pixel.

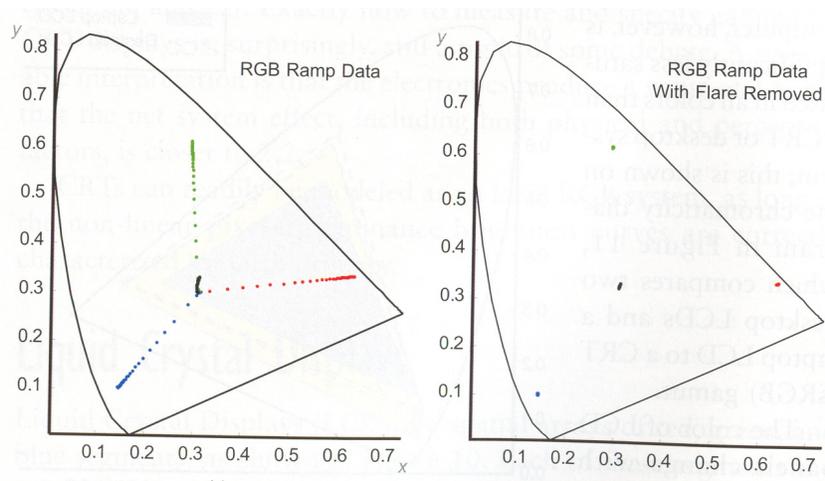


Figure 6.8: The effects of flare light on the chromaticity of the primaries of a LCD display. Left: With flare. Right: With flare subtracted.

adds a constant level of light to all the primaries rather than having them be constant for all intensities. Thus, the chromaticities of the primaries shift towards the white point for the lower intensities. Even with flare subtracted, the LCDs deviate measurably from the ideal RGB model since the chromaticity is still not constant at all brightness levels. This is illustrated in Figure 1.8.

6.1.7 Projection Displays

A digital projector contains a digital imaging element like an LCD panel or an array of digital micromirrors that modulate the light from a high intensity light bulb. Most LCD projectors and the larger DMD projectors have three imaging elements and a dichroic mirror that splits the white light from the bulb into its red, green and blue components. These are recombined and displayed through a single lens. The smaller projectors use a single imaging element with a wheel of filters, so that the separations are displayed sequentially in time. Some DLP (digital light processing) projectors have a fourth filter called the clear filter which are used while projecting the grays to achieve a higher brightness for grays. Also, note that this is equivalent to using more than three primaries where of course we can reproduce more colors. However, the advantage of having a three primary system is that it

forms a coordinate system with two linearly independent coordinate axes (two sides of the triangle). Thus defining any color in terms of this coordinate is *unique* and every color can be generated from the primaries in an unique fashion. However, when four primaries are used, this is no longer true. The primaries can be mixed in more than one ways to generate a color. This brings in ambiguities in a device that is difficult to handle. The color for a projection system is thus guided by the color of its bulb and the filters. However, the primaries for projectors are often significantly different from those of the sRGB standard. Projectors being of LCD or DMD technology also exhibit the black offset problem.

6.2 Subtractive Devices

Subtractive color systems put colored images in paper, either by printing with ink or by exposing special photographic paper. Subtractive color uses layers of cyan, magenta and yellow filters to create color. In photography, the dye layers of the film sequentially absorb the blue, green and red light. When developed, white light filtered through the layers recreates the image. Each layer modulates one of the red, green or blue component of the white light, leaving the other components unchanged. These layers plus the white light defines the subtractive color system. Subtractive color depends on selective absorption and hence depends critically on the spectral distribution of the illuminant.

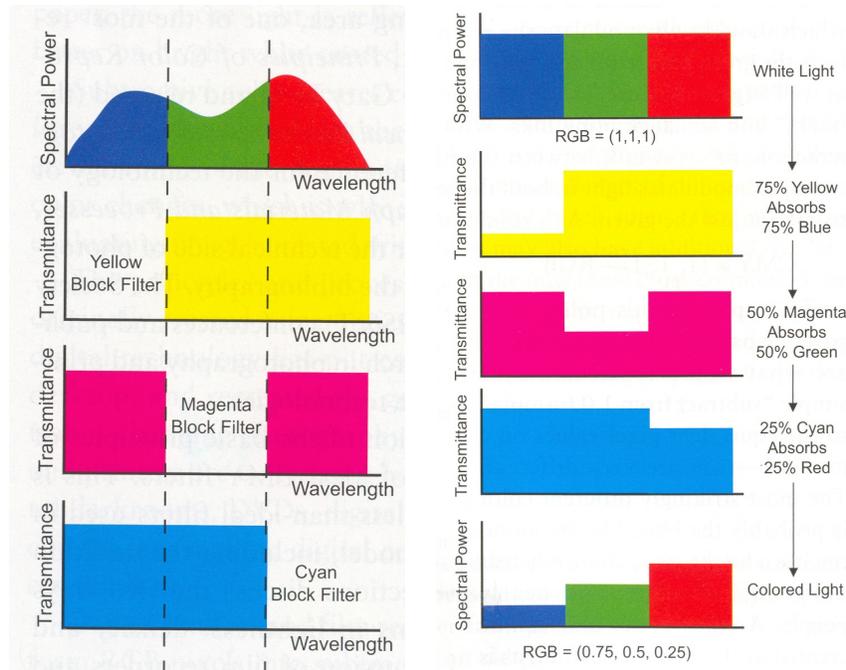


Figure 6.9: Left: Ideal cyan, magenta and yellow block filters. Right: Color formation using block filter primaries.

The inks and dyes in the subtractive color system act as absorption filters. The filter that passes green and blue but modulates red, appears cyan. Similarly, the green-modulating filter appears magenta and the blue-modulating filter appears yellow. Ideal form of these filters, called block filters, are shown in Figure 1.9. Overlaying all of these primary colors lead to the absorption of all wavelengths creating black (instead of white in the additive color system).

Figure 1.9 shows how to create the color ($R : 0.75, G : 0.5, B : 0.25$) using ideal block filters. White light enters from top. The yellow filter subtracts 75% of blue. Next, the magenta filter subtracts 50% of green and then

the cyan filter subtracts 25% of red. The transmitted light is $(R : 0.75, G : 0.5, B : 0.25)$. Thus, in the case of ideal block filters, the absorption of the filters is the inverse of the RGB intensity of light. That is,

$$(C : 0.25, M : 0.5, Y : 0.75) = (1, 1, 1) - (R : 0.75, G : 0.5, B : 0.25)$$

But it is important to emphasize that this simple model works only for ideal block filters.

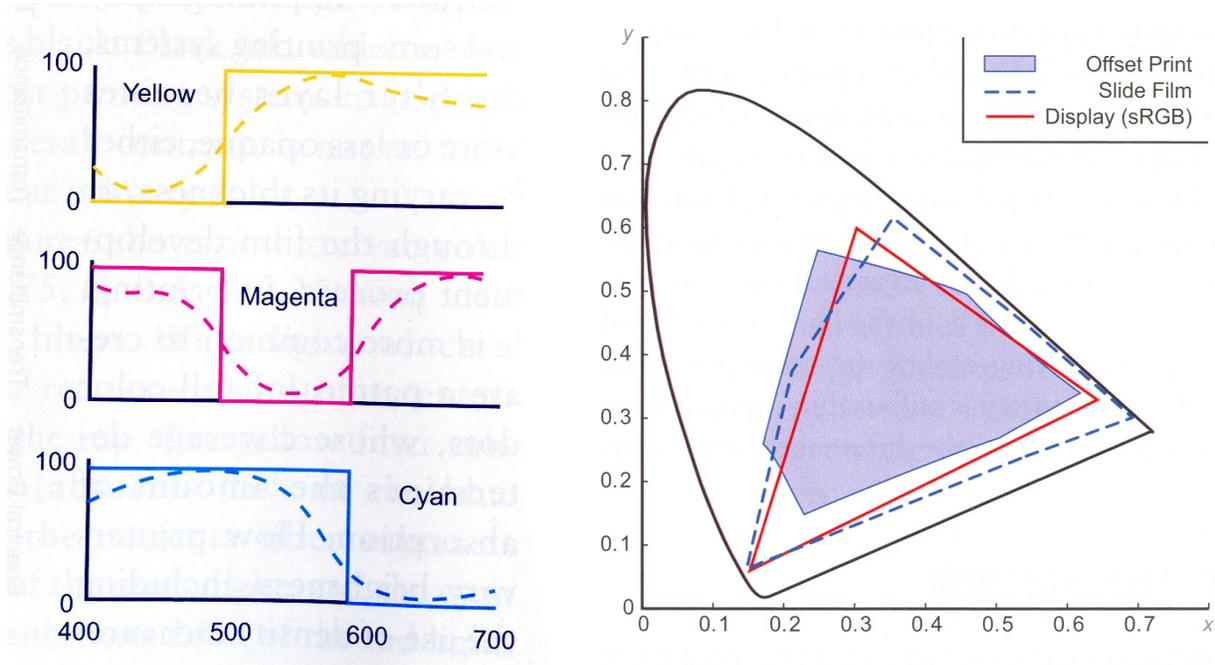


Figure 6.10: Left: Real cyan, magenta and yellow block filters. Right: Comparing additive gamuts with subtractive gamuts.

However, the real subtractive filters are never ideal. The cyan filter which should only affect the red part of the spectrum, also affects the green part. Figure 1.10 shows a comparison of real-life filters with the ideal block filters. Thus in real filters, there are interactions or cross talk amongst the filters. As a result,

$$CMY \neq (1, 1, 1) - RGB$$

This is mainly due to ink impurities. The effect of ink impurities is especially pronounced in the reproduction of gray. With ideal block filters, equal amount of primary should create a neutral gray. But that hardly happens in real filters. The different filters need to be balanced in different ways to create different levels of gray, which makes reproduction of the neutral gray scale extremely difficult with subtractive color.

In printing, black is mainly used to improve contrast. It is also used to replace expensive colored ink with cheaper black ink, and to prevent the paper from getting too wet. Black in this context is called K, for *key*, an old printing term. Note that black is not an independent primary, hence it can only increase the gamut by making the dark colors darker. For any color that has all the three components, there is a *gray component* equal to the minimum of the three. For example, in $(C : 0.5, M : 0.5, Y : 0.25)$, the gray component is 0.25. This gray component is the maximum amount of black that can be used while reproducing this color. Determining the amount of black to be applied and the amount of cyan, magenta and yellow dyes to be reduced is called *gray component replacement* or GCR. Previously black was used only for the reproduction of gray scale and was called *undercolor removal* or UCR. Full GCR would replace the whole gray component with black. But in reality full GCR are often not used.

Using four colors instead often achieves better contrast. However, often due to registration problems, especially in inexpensive inkjet printers, full or almost full GCR is used.

The process of defining black component in printing starts with creating a neutral gray scale by a blend of the four different types of ink. Note that a different combination may be used for all the different levels of gray in the gray axis. This process is called *gray balancing*. Usually black is added to the mix starting at around 50%. However note that there is no mathematics to predict how to balance the CMY to create a neutral gray - it depends completely on the type of paper, ink and process used. The traditional solution was to plot a large number of patches with nearly equal CMY values, search for grayest one, and then iterate until the good grayscale is found.

Subtractive color reproduction is a means of taking an image encoded as RGB pixels and reproducing it on subtractive color system. In terms of device independent color reproduction, it means mapping from the additive gamut of the input pixels to the subtractive gamut of the output system. Subtractive color gamuts are very different from additive gamuts as illustrated in Figure 1.10. Even this figure understates the problem, since many colors that appear to fall within both the gamut may not actually do so due to the difference in the brightness range. This will be dealt in details in the next section.

Note that there is no simple tristimulus model for subtractive color. The model for subtractive color depends on specific spectral distributions, including that of ambient light, not just tristimulus values. It is also dependent on a larger set of variables like reflectance properties of the paper, exposure times, ink consistency etc.

6.3 Gamut Matching

However, note that when going across different devices that have different color gamut, tone mapping and dynamic range, it is important to do a correct conversion to retain the image appearance close to the original. For example, if we have a second device whose color gamut is denoted by the orange triangle in Figure 1.2, each of the devices have some colors that cannot be represented by the other. Only the colors in the intersection of two gamuts can be represented by both devices. This is the common problem of gamut matching faced while matching colors across printers, monitors and other devices like cameras or projectors.

For additive color displays, all the parameters defining the color gamut (explained in Section 1.1.1) can be derived from the XYZ values of the three primaries and the black (for displays where black offset is evident). These define a rectangular parallelepiped in the XYZ space that defines the 3D color gamut as illustrated in Figure 1.2. The tone mapping function defines the parameterization of each of the axis of this parallelepiped. Thus, changing from the gamut of one display to another is a 4x4 linear transformation.

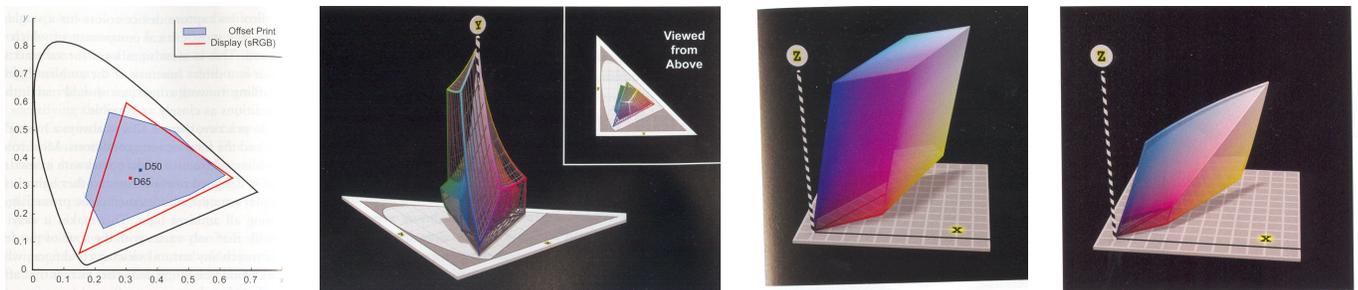


Figure 6.11: Left to Right: An additive and subtractive gamut on the chromaticity diagram, the same gamuts plotted in xyY space to show the difference in luminance values as well as chromaticity (viewed from above it reduces to the left figure), the additive gamut in XYZ space, the subtractive gamut in XYZ space.

It is important to note here that a true comparison of gamuts can only happen in 3D, a point often lost in the convenient use of the chromaticity coordinates. This is especially true when we are trying to match the gamut

of an additive color system to that of a subtractive color system, as illustrated in Figure 1.11. The same color gamuts plotted on the chromaticity diagram when plotted in the xyY space shows that the gamuts rise above the surface of the chromaticity diagram in proportion to their luminance. The wireframe shows the additive monitor color gamut while the solid one shows the subtractive printer gamut. Thus, though in 2D one seems to contain the other, it is not true in 3D. Also, Figure 1.11 also shows the two gamuts plotted in XYZ color space. Note that the additive gamut forms a nice parallelepiped indicating a linear transformation between the XYZ and this RGB space. However, the print gamut is distinctly non-linear; while there is some hint of the original color cube, each face has different size and some of the edges are distinctly curved. Thus, the gamut is actually concave and bent.

Most commercial color management systems match the gamut in CIE LAB space. First the neutral axis of both the gamuts are aligned. Then the out-of-gamut colors are projected toward the center of the gamut in a way that reduces saturation, and to a lesser extent brightness, while maintaining hue. This is called *gamut mapping*. No single gamut mapping method is often adequate for all images and different methods can be used depending on the application. Following are some popular gamut mapping methods.

1. *Perceptual*: Create an aesthetic mapping for images. Maps all colors smoothly into the target gamut, though those outside the gamut will move relatively more than those inside the gamut.
2. *Saturated*: This is usually used for charts and illustrative graphics that contains solid saturated colors. This mapping maintains the relative saturation while sacrificing the smoothness.
3. *Relative colorimetric*: This uses minimal color transformation that projects the out-of-gamut colors to the gamut surface with little or no-mapping of in-gamut colors. This is the default mapping for older printing systems.