
Digital Image Basics

Written by Jonathan Sachs
Copyright © 1996-1999 Digital Light & Color

Introduction

When using digital equipment to capture, store, modify and view photographic images, they must first be converted to a set of numbers in a process called digitization or scanning. Computers are very good at storing and manipulating numbers, so once your image has been digitized you can use your computer to archive, examine, alter, display, transmit, or print your photographs in an incredible variety of ways.

Pixels and Bitmaps

Digital images are composed of *pixels* (short for picture elements). Each pixel represents the color (or gray level for black and white photos) at a single point in the image, so a pixel is like a tiny dot of a particular color. By measuring the color of an image at a large number of points, we can create a digital approximation of the image from which a copy of the original can be reconstructed. Pixels are a little like grain particles in a conventional photographic image, but arranged in a regular pattern of rows and columns and store information somewhat differently. A digital image is a rectangular array of pixels sometimes called a *bitmap*.

Types of Digital Images

For photographic purposes, there are two important types of digital images—color and black and white. Color images are made up of colored pixels while black and white images are made of pixels in different shades of gray.

Black and White Images

A black and white image is made up of pixels each of which holds a single number corresponding to the gray level of the image at a particular location. These gray levels span the full range from black to white in a series of very fine steps, normally 256 different grays. Since the eye can barely distinguish about 200 different gray levels, this is enough to give the illusion of a stepless tonal scale as illustrated below:



Assuming 256 gray levels, each black and white pixel can be stored in a single byte (8 bits) of memory.

Color Images

A color image is made up of pixels each of which holds three numbers corresponding to the red, green, and blue levels of the image at a particular location. Red, green, and blue (sometimes referred to as RGB) are the primary colors for mixing light—these so-called additive primary colors are different from the subtractive primary colors used for mixing paints (cyan, magenta, and yellow). Any color can be created by mixing the correct amounts of red, green, and blue light. Assuming 256 levels for each primary, each color pixel can be stored in three bytes (24 bits) of memory. This corresponds to roughly 16.7 million different possible colors.

Note that for images of the same size, a black and white version will use three times less memory than a color version.

Binary or Bilevel Images

Binary images use only a single bit to represent each pixel. Since a bit can only exist in two states—on or off, every pixel in a binary image must be one of two colors, usually black or white. This inability to represent intermediate shades of gray is what limits their usefulness in dealing with photographic images.

Indexed Color Images

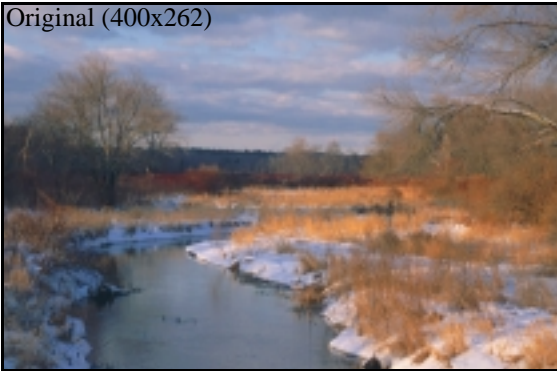
Some color images are created using a limited palette of colors, typically 256 different colors. These images are referred to as indexed color images because the data for each pixel consists of a palette index indicating which of the colors in the palette applies to that pixel. There are several problems with using indexed color to represent photographic images. First, if the image contains more different colors than are in the palette, techniques such as dithering must be applied to represent the missing colors and this degrades the image. Second, combining two indexed color images that use different palettes or even retouching part of a single indexed color image creates problems because of the limited number of available colors.

Resolution

The more points at which we sample the image by measuring its color, the more detail we can capture. The density of pixels in an image is referred to as its *resolution*. The higher the resolution, the more information the image contains. If we keep the image size the same and increase the resolution, the image gets sharper and more detailed. Alternatively, with a higher resolution image, we can produce a larger image with the same amount of detail.

For example, the following images illustrate what happens as we reduce the resolution of an image while keeping its size the same—the pixels get larger and larger and there is less and less detail in the image:

Original (400x262)



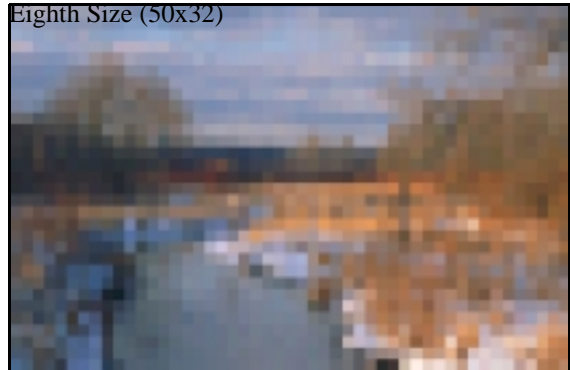
Half Size (200x131)



Quarter Size (100x65)



Eighth Size (50x32)



As we reduce the resolution of an image while keeping its pixels the same size—the image gets smaller and smaller while the amount of detail (per square inch) stays the same:



Color Terminology

While pixels are normally stored within the computer according to their red, green, and blue levels, this method of specifying colors (sometimes called the RGB *color space*) does not correspond to the way we normally perceive and categorize colors. There are many different ways to specify colors, but the most useful ones work by separating out the hue, saturation, and brightness components of a color.

Primary and Secondary Colors and Additive and Subtractive Color Mixing

Primary colors are those that cannot be created by mixing other colors. Because of the way we perceive colors using three different sets of wavelengths, there are three primary colors. Any color can be represented as some mixture of these three primary colors.

There are two different ways to combine colors—*additive* and *subtractive* color mixing.

Subtractive color mixing is the one most of us learned in school, and it describes how two colored paints or inks combine on a piece of paper. The three subtractive

primaries are Cyan (blue-green), Magenta (purple-red), and Yellow (not Blue, Red, and Yellow as we were taught).

Additive color mixing refers to combining lights of two different colors, for example by shining two colored spotlights on the same white wall. The additive color model is the one used in computer displays as the image is formed on the face of the monitor by combining beams of red, green, and blue light in different proportions.

Color printers use the subtractive color model and use cyan, magenta, and yellow inks. To compensate for the impure nature of most printing inks, a fourth color, black is also used since the black obtained by combining cyan, magenta, and yellow inks is often a murky dark green rather than a deep, rich black. For this and other reasons, commercial color printing presses use a 4-color process to reproduce color images in magazines.

A color created by mixing equal amounts of two primary colors is called a *secondary*. In the additive color system, the primary colors are:



and the secondary colors are:



In the subtractive color system, the roles of the primaries and secondaries are reversed.

Color Gamut

In the real world, the ideal of creating any visible color by mixing three primary colors is never actually achieved. The dyes, pigments, and phosphors used to create colors on paper or computer screens are imperfect and cannot recreate the full range of visible colors. The actual range of colors achievable by a particular device or medium is called its color gamut and this is mostly but not entirely determined by the characteristics of its primary colors. Since different devices such as computer monitors, printers, scanners, and photographic film all have different color gamuts, the problem of achieving consistent color is quite challenging. Different media also differ in their total dynamic range—how dark is the darkest achievable black and how light is the brightest white.

Color Management

The process of getting an image to look the same between two or more different media or devices is called *color management*, and there are many different color management systems available today. Unfortunately, most are complex, expensive, and not available for a full range of devices.

Hue

The hue of a color identifies what is commonly called “color”. For example, all reds have a similar hue value whether they are light, dark, intense, or pastel.

Saturation

The saturation of a color identifies how pure or intense the color is. A fully saturated color is deep and brilliant—as the saturation decreases, the color gets paler and more washed out until it eventually fades to neutral.

Brightness

The brightness of a color identifies how light or dark the color is. Any color whose brightness is zero is black, regardless of its hue or saturation. There are different schemes for specifying a color's brightness and depending on which one is used, the results of lightening a color can vary considerably.

Luminance

The luminance of a color is a measure of its perceived brightness. The computation of luminance takes into account the fact that the human eye is far more sensitive to certain colors (like yellow-green) than to others (like blue).

Chrominance

Chrominance is a complementary concept to luminance. If you think of how a television signal works, there are two components—a black and white image which represents the luminance and a color signal which contains the chrominance information. Chrominance is a 2-dimensional color space that represents hue and saturation, independent of brightness.

Viewing Images on a Monitor

Before you can get consistent results working with digital images, you must create a standard viewing environment.

Most computer monitors use a standard set of *phosphors* (the chemical compounds on the face of the screen that are responsible for creating its primary colors). These phosphors, combined with the electronic circuitry that activates them create color images on your computer screen. To get an accurate match between what you see on your monitor and prints or other output requires that a number of variables be controlled:

Viewing Conditions

First, it is very important for critical evaluation of photographic images on a computer monitor to work in an environment with subdued lighting. When strong room light falls on the face of the monitor, it makes it almost impossible to discern shadow detail and seriously alters the appearance of many of the colors.

Monitor Brightness Control

Next, you must set the brightness control on your monitor correctly. If the level is too high, blacks will start to become grays and colors will be washed out. If the level is too low, there will be loss of shadow detail and colors will appear murky.

Gamma

Since the human eye has a nonlinear response curve (it is more sensitive to variations in low light than to an equal variation in bright light), it is important to track this nonlinearity with the display of different gray levels. The curve that relates the amount of light emitted from the monitor to the 256 different gray levels in the digital image is called the *gamma* curve. There are several different standard gamma curves in common use. They all follow a power law of the following form:

$$b = v^\gamma$$

where b is the brightness of the light emitted from the screen, v is the gray level value in the image and γ is the monitor gamma. The standard value for γ that most closely approximates the response of the human eye is 2.22. This is the value used by the television industry for storing and displaying video images. The prepress

industry has standardized on a value of 1.8 which more closely corresponds to the characteristics of printing presses. Macintosh computers often use the value 1.4 for some unknown reason, while most uncorrected PC monitors have a gamma of about 2.3.

If an image was designed to be displayed on a monitor with a gamma of 2.2 and the monitor's actual gamma is set to 1.8, the image will appear too light and washed out. Conversely, if an image was adjusted to look good using a computer with a monitor gamma of 1.4 or 1.8 and viewed at 2.2, it will look too dark. For this reason, it is very important to set your computer's monitor to the same gamma setting as will be used to print it. If gamma is ignored, there is no way to guarantee that prints will not be too light or too dark.

White Point

Finally, there is an important characteristic of computer monitors called *white point*. This is usually given as a color temperature and for most uncorrected monitors, it runs about 9300 degrees. What this means is that the white that is displayed on the monitor when you send it maximum values of red, green, and blue has a distinctly bluish cast. If you were going to view your prints under a light of this same color temperature, you would get a good match (all other variables being set properly). Different lights like incandescent, quartz halogen, fluorescent, and daylight (morning, evening, indirect, etc.) all have different color temperatures and prints viewed using these different lighting conditions will look radically different. The standard white point used in the graphic arts industry is a 5000 degree color temperature and most light boxes are set to roughly this color. Viewing prints under fluorescent lighting can be particularly deceptive as mercury vapor lamps have several spikes in the green part of the spectrum whereas daylight has a relatively smooth spectrum. This can give rise to considerable variation in certain colors.

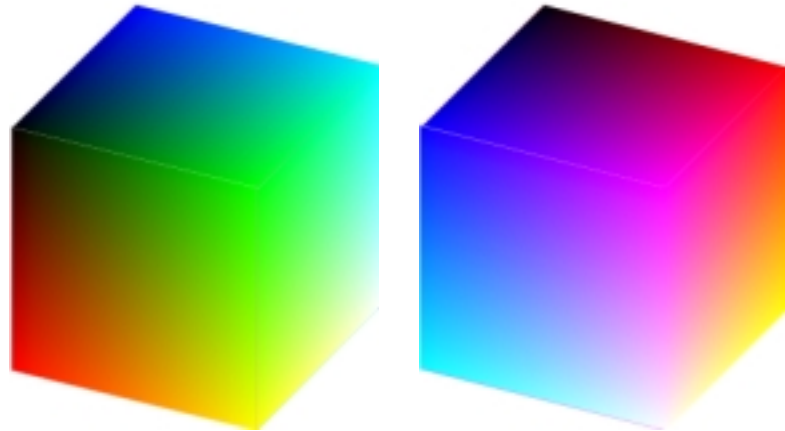
Two monitors which use the same phosphors, are both viewed in subdued light, have their gamma set the same, have their brightness control set properly, and have the same white point, should produce nearly identical displays.

Color Spaces

A color space is a mathematical system for representing colors. Since it takes at least three independent measurements to determine a color, most color spaces are three-dimensional. Many different color spaces have been created over the years in an effort to categorize the full gamut of possible colors according to different characteristics. Picture Window uses three different color spaces:

RGB

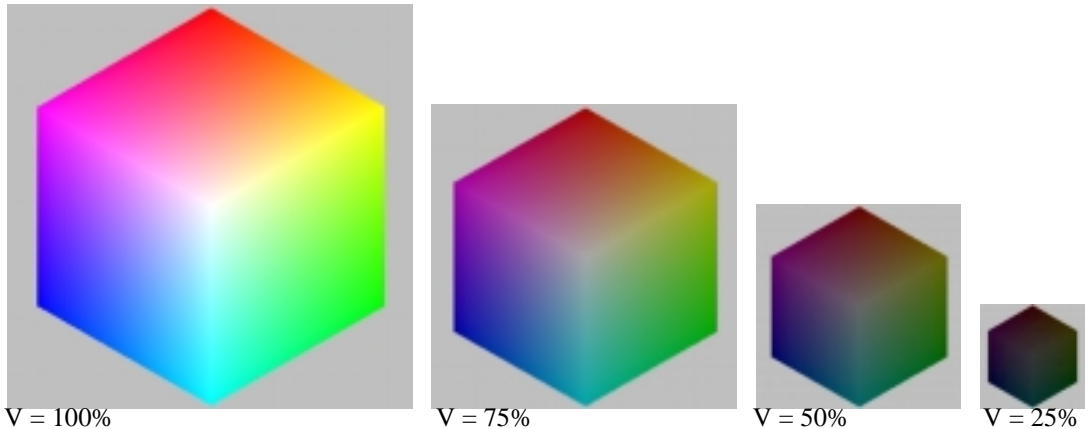
Most computer monitors work by specifying colors according to their red, green, and blue components. These three values define a 3-dimensional color space call the RGB color space. The RGB color space can be visualized as a cube with red varying along one axis, green varying along the second, and blue varying along the third. Every color that can be created by mixing red, green, and blue light is located somewhere within the cube. The following images show the outside of the RGB cube viewed from two different directions:



The eight corners of the cube correspond to the three primary colors (Red, Green, Blue), the three secondary colors (Cyan, Magenta, Yellow) and black and white. All the different neutral grays are located on the diagonal of the cube that connects the black and the white vertices.

HSV (Hue Saturation Value)

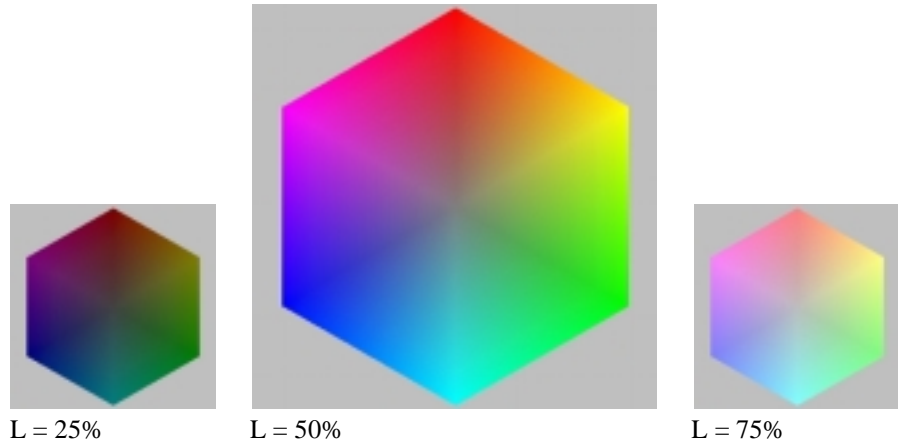
The HSV color space attempts to characterize colors according to their hue, saturation, and value (brightness). This color space is based on a so-called hexcone model which can be visualized as a prism with a hexagon on one end that tapers down to a single point at the other. The hexagonal face of the prism is derived by looking at the RGB cube centered on its white corner. The cube, when viewed from this angle, looks like a hexagon with white in the center and the primary and secondary colors making up the six vertices of the hexagon. This color hexagon is the one Picture Window uses in its color picker to display the brightest possible versions of all possible colors based on their hue and saturation. Successive crosssections of the HSV hexcone as it narrows to its vertex are illustrated below showing how the colors get darker and darker, eventually reaching black.



HSL (Hue Saturation Lightness)

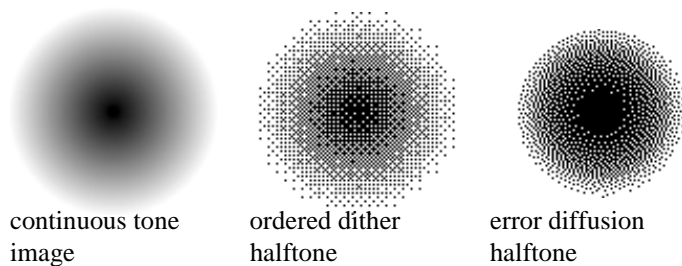
The HSL color space (also sometimes called HSB) attempts to characterize colors according to their hue, saturation, and lightness (brightness). This color space is based on a double hexcone model which consists of a hexagon in the middle that converges down to a point at each end. Like the HSV color space, the HSL space

goes to black at one end, but unlike HSV, it tends toward white at the opposite end. The most saturated colors appear in the middle. Note that unlike in the HSL color space, this central crosssection has 50% gray in the center and not white.



Continuous Tone vs. Halftone Images

There are many different technologies used to reproduce photographic images. Some printing technologies such as laser and ink jet printers work by creating individual dots, each of which is one of a small number of solid colors. For example, a black and white laser printer can only print black or white pixels -- it cannot produce gray dots. To create the illusion of a photographic image, these tiny dots must be clustered in different proportions to reproduce the different colors and gray levels in the image. This dot clustering process is called *halftoning* or *dithering* and it can be done in many different ways.



Devices such as computer monitors, dye sublimation printers, and film recorders can create dots of any color without halftoning, and these are called *continuous tone* printers because they can reproduce a continuous range of gray levels or colors.

Because a continuous tone printer packs more information into each dot, to achieve comparable image quality, printers that use halftoning or dithering require higher resolution, by a factor of roughly 6-8. Thus a continuous tone printer that outputs images at 200 dpi is roughly similar in quality to a halftone printer that outputs pixels at roughly 1200-1600 dpi. For this reason, even a 600 dpi laser printer makes relatively coarse images equivalent to only about 75 dpi continuous tone output.

For printed output to appear photographic when viewed from a standard distance of about 10 inches requires between 150 and 300 dpi. At 150 dpi, most images will appear noticeable soft, while between 200 and 300 dpi the incremental improvement is quite small. Beyond 300 dpi there is still some improvement, but for most images and most observers, it will be virtually unnoticeable.

How images are digitized

The process of converting an image to pixels is called *digitizing* or *scanning* and this function can be performed in many different ways.

Film Scanners

This type of scanner is sometimes called a slide or transparency scanner. They are specifically designed for scanning film, usually 35mm slides or negatives, but some of the more expensive ones can also scan medium and large format film. These scanners work by passing a narrowly focused beam of light through the film and reading the intensity and color of the light that emerges.

Flatbed Scanners

This type of scanner is sometimes called a reflective scanner. They are designed for scanning prints or other flat, opaque materials. These scanners work by shining white light onto the object and reading the intensity and color of the light that is reflected from it, usually a line at a time. Some flatbed scanners have available transparency adapters, but for a number of reasons, in most cases these are not very well suited to scanning film. On the other hand, flatbed scanners can be used as a sort of lensless camera to directly digitize flat objects like leaves.

Digital Cameras

One of the most direct ways to capture an image is a digital camera which uses a special semiconductor chip called a CCD (charge coupled device) to convert light to electrical signals right at the image plane. The quality of the images created in this manner is closely related to the number of pixels the CCD can capture. Affordable digital cameras suffer from relatively low resolution, limited dynamic range, and low ISO film speed equivalent, and consequently do not always produce high quality digital images. To get images with quality comparable to film photography currently requires very expensive digital cameras.

Video Frame Grabbers

This type of scanner uses a video camera to capture a scene or object and then converts the video signal that comes out of the camera to a digital image in your computer memory. A video camera can be used to digitize scenes containing 3-dimensional objects, but they usually have much lower image quality than film or flatbed scanners.

Scanning Services

Photo CD is a service started by Kodak a number of years ago whereby your film can be scanned using a high quality scanner and written to a compact disk you computer can access. Using Photo CD service is an inexpensive way to get high quality scans of your images without purchasing a scanner.

Many other scanning services are available which can scan prints or film to floppy disks or removable disk cartridges. These vary from low resolution snapshot quality images to professional drum scans at very high resolution.