

2a

Introduction to Digital Photography

The proverbial handwriting isn't just on the wall, it's emblazoned in neon lights that you'd have to be blind to miss: digital cameras will unquestionably take over from cameras that depend on film, just as the digital technology of CDs took over from vinyl records, and just as digital cell phones are taking over from analog cell phones. It may not happen quickly, but it will happen.

Digital Cameras are Here to Stay

Monochrome generally refers to black and white images. However, there are monochrome monitors that display only one foreground and one background color, such as what you see when you open a command prompt on your computer.

If we had to pick two major trends to bet on as most likely to take over in any technology that they apply to at all, the choice would be easy. The first half of our bet would be on the trend toward color. The second half would be on the trend toward digital technologies. As it happens, both are critical elements for digital photography, and both trends have already taken over as much as they need to for digital photography to flourish.

Color first: the human brain seems hard-wired to prefer color. Movies, still photographs, televisions, computer screens, and more all started out in black and white—or, more precisely, *monochrome*—formats, and all have moved (mostly) to color.

From a photographer's point of view—and that includes everyone from professionals to the most occasional amateur—the two most important technologies that have switched to color are computer monitors and computer printers. And this is

Having a good quality monitor and printer is only half the battle. Both must be calibrated so that what you see on the screen is what is printed at the printer.

Most high-end, third-party photo-editing software and printers come with their own calibration software.

a case where the benefit of having color available in both technologies is much greater than the sum of the parts.

Having good color quality on both monitors and printers opens up more possibilities. Not only can you do everything that you could do with a color monitor alone plus what you could do with a color printer alone, but you can do more. The combination gives you what amounts to your own sophisticated photo lab tucked inside your computer. Take a picture, and you can print it, e-mail it, or post it on a Web site as is. Or you can enhance it, add graphic elements or text, or crop it (cut off part of the image) and resize it, as in Figure 2-1, and *then* print it, e-mail it, or post it.

Note The photo samples in these lessons will all be monochrome. Little is lost, however, because there is little we would be able to show in color that we can't show effectively in monochrome format as well.



Figure 2-1.

The top right picture and the two bottom pictures are all edited versions of the top left original.

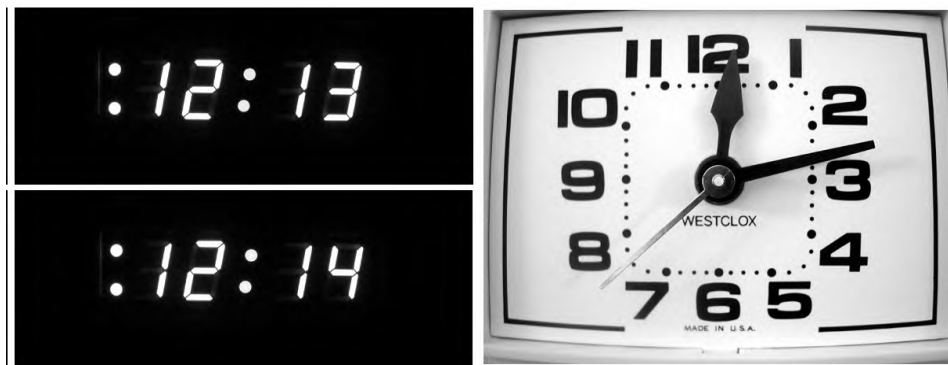
For a professional photographer dealing in, say, portraits or corporate brochures or annual reports, this ability to quickly and easily modify an image on his or her computer screen not only saves time and money, but gives better control over the final product.

For a nonprofessional, it opens the door to doing things that you would never have done with film—no, make that never would have even *thought* of doing with film. Before we start the camera clicking, let's learn a bit about how digital technology works.

A Bit about the Digital World

The driving force pushing us all to digital technologies comes from a very different direction than the drive to color. In truth, people's perceptions aren't wired for digital anything the way they are wired for color. To the contrary, they are wired for *analog* technologies, which allow continuous change rather than the discrete steps that digital technologies offer.

Consider a digital clock versus an analog clock as shown in Figure 2-2, for example. The digital clock shows hours and minutes. It can show you when the time is, say, 12:13 or 12:14, but it can't show anything in between. Analog clocks—by which we mean real analog clocks, preferably with a sweep second hand, not the recreations drawn on the face of some digital clocks—can't skip the in-between times. To get from 12:13 to 12:14, the second hand has to trace out all the seconds in between.



The human voice is an analog signal, so phone lines carry analog data.

Computers, on the other hand, transmit data only digitally. When data is transmitted over a phone line, a computer modem is used to convert the incoming analog signal to a digital one; then convert it back for outgoing transmissions.

Figure 2-2.

Digital clocks show time in discrete steps, jumping from one second to the next. Analog clocks show time as falling along a continuous circle.

That's the essential difference between digital and analog technologies, and it's really all you need to know about the differences: Digital technology changes things in steps. Analog technology changes things continuously.

This difference is so obvious and so simple that it's easy to miss how important it is and in how many different ways it comes into play. Another example or two will help. (We've made a point of picking examples that relate directly to digital photography and will give you a basic understanding of some of its possibilities and weaknesses.)

We'll start with a simple case. Suppose that you want to draw a diagonal line on a piece of paper, going down, from left to right. The technology you're using—pencil

and paper—is analog in nature. (You probably don’t usually think of pencil and paper as technology, but it is. Consider how much trouble it would be to run a civilization without it.) That means you can drag the pencil point along the paper—preferably using a ruler to get a straight line—and wind up with a nice, smooth diagonal line, like the one in Figure 2-3.

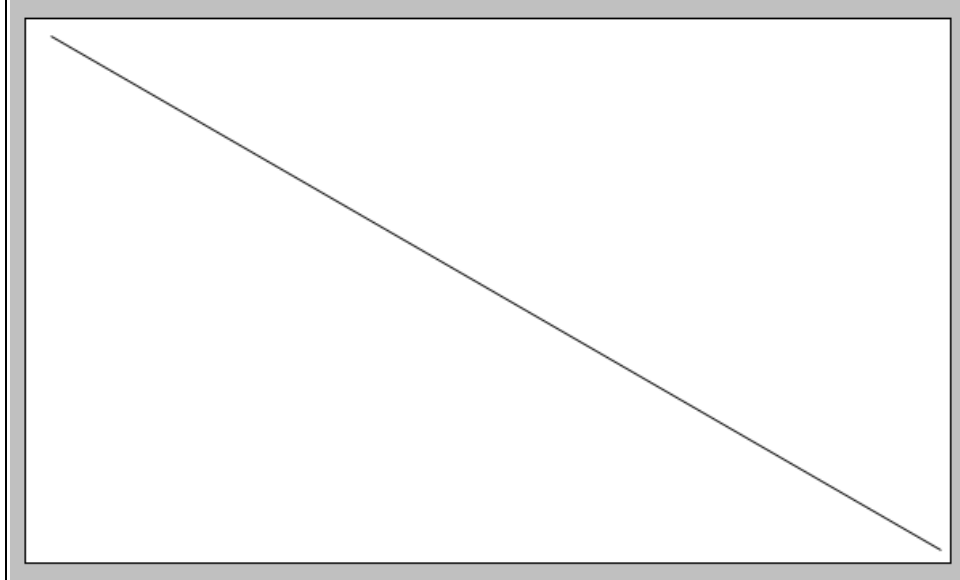


Figure 2-3.

An analog technology like a pencil will give you a smooth line, like this one.

If you try the same thing with a digital technology, drawing a line with a drawing program on a computer monitor for example, the line comes out in steps—one step right, one step down, one step right, and so forth, like the one in Figure 2-4. The result is anything but smooth.

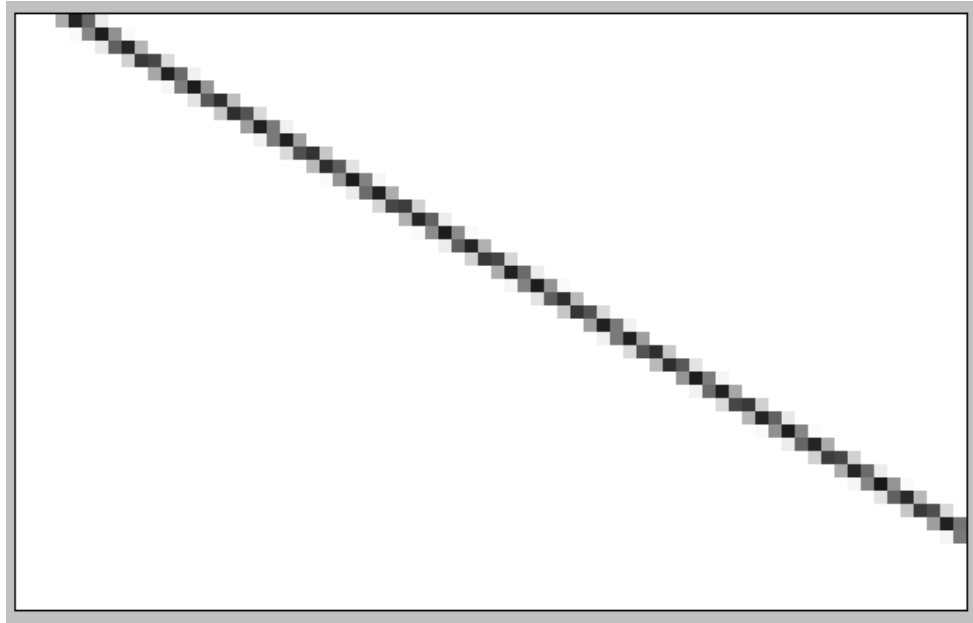


Figure 2-4.

A digital technology will give you a staircase step effect for a diagonal line.

The steps that use lighter grays in this line are meant to help blur the steps and give the line a smoother looking edge than if it used black only. Even with those blocks of light gray, however, this second example of a diagonal line is not an acceptable substitute for a straight line—at least, not to the human visual system. But now comes a subtle irony.

In the real world, at the most basic physical level, everything comes in steps. The graphite in the line laid down by the pencil and, for that matter, the paper itself, both have a built-in graininess. More than that, they are grainy on all sorts of levels—from a roughness in the paper surface visible even with a good magnifying glass, to individual molecules and atoms, and even to levels below that. (And that's as far as we will venture into physics).

The point is that if the steps are small enough, you can't see them. Or, to put it another way: if a digital technology breaks the steps down far enough, it can effectively erase the difference between digital and analog technologies. As proof of that point, note that Figures 2-3 and 2-4 both show the same line drawn in the same program. The difference is that we zoomed in on the line to get the second figure and make the steps far larger and more visible.

Digital Cameras and Resolution

A megapixel is 1 million pixels. An image with 6 million pixels offers a better quality image for a larger print than one that offers only 2 million pixels because the pixels are smaller and less noticeable.

A pixel is the smallest building block of color on a computer display of a digital image.

Probably the most obvious area where the idea of digital steps comes into play for digital cameras is resolution—the number of pixels the camera uses. If you know anything about digital cameras at all, it's probably that they are rated in megapixels and that more megapixels is better. A 3-megapixel camera offers better pictures than a 2-megapixel camera, and a 6-megapixel camera offers better pictures still. More megapixels also makes for a more expensive camera. In case you're not too clear on what a megapixel is, however, we'll start with some basics.

Digital photos are similar to mosaics, which create pictures from the arrangement of small tiles. In a mosaic, the tile is the smallest unit of the picture. In a digital photo, you use *picture elements*, usually abbreviated as *pixels*, and the pixel is the smallest unit of the picture. Whether you're dealing with tiles or pixels, each one has a specific color, and it's the arrangement of the tiles or pixels that creates the picture.

A *megapixel* is a million pixels, more or less. (Actual cameras tend to round their claims to the nearest megapixel, or sometimes the nearest tenth of a megapixel.) The exact numbers aren't important. What you need to know for the moment is that the pixels are arranged in a rectangular area—so many pixels across by so many pixels down. Multiply the number across by the number down, and you'll wind up with the maximum number of pixels that the camera delivers. A camera that shoots photos at 1600×1200 resolution, for example, delivers 1,920,000 pixels (and would usually be called a 2-megapixel camera).

The important point is that any given camera, and, therefore, any given photo the camera takes, is limited to some maximum number of pixels. That number stays the same whether you print the picture at 3×4 inches, 8×10 inches, or big enough to fill a billboard. It also stays the same whether you show the photo on a monitor with 640×480 resolution, or expand it to fill a monitor with 1600×1200 resolution.

The potential problem—which is where the issue of steps comes in—is that no matter what size you print the photo or expand it to on screen, the size of the pixels stays the same relative to the full photo size. (This is a bit of an oversimplification, because there are ways around the issue, but it's essentially right.) Print the picture at a large enough size, or expand it to a large enough size on screen, and the photo may get *pixelized*, so you can see the individual pixels.

Figures 2-5 through 2-7 show the same photo as it looks on screen at different levels of zoom. In Figure 2-5 the photo is set up at 100 percent zoom, which means each pixel on the screen uses one pixel in the photo. There's no pixelization at all.



Figure 2-5.

A photo as it appears on screen at 100 percent, using one pixel on screen for each pixel in the photo.

In Figure 2-6, we zoomed in to 200 percent, which is equivalent to taking a photo that looks fine at 3 × 5 inches and enlarging it to 6 × 10 inches—or taking a photo with 640 × 480 resolution and enlarging it to fill a monitor screen running at 1280 × 1024 resolution. By doubling both the height and width of the image, you can begin to see a hint of the pixels, particularly in the whiskers in the lower left of the picture.



Figure 2-6.

The same photo as it appears at 200 percent, showing some hints of pixelization, particularly in the whiskers at the bottom left.

In Figure 2-7, finally, we zoomed in to 400 percent, doubling the size of the picture again. This is equivalent to taking a picture designed for printing at 3×5 inches, deciding to focus on an area about half that height and width, and then printing it at 8×10 inches. Here you can clearly see the small rectangular pixels in the eye and in the cat's surrounding fur.



Figure 2-7.

At 400 percent the pixelization in this photo is much more obvious.

Once again, the basic point here is that as long as the steps are small enough, as in Figure 2-5, you won't see them. As the steps get larger, they get more visible. If we had gone to still greater enlargement than in Figure 2-7, the pixels would be even more obvious.

About Color

Digitizing is important to many different industries. For instance, optical scanners digitize scanned images for artists, and libraries digitize books so they can be read at the computer.

The idea that if you use small enough steps you can effectively provide continuous, analog-like results applies to color as well. As we've already pointed out, color is one of the most important features in digital photography. Without it, digital cameras could take photos with as many pixels as you'd ever need, and they'd still have limited appeal. Color is also one of the features that gets *digitized*, or converted into digital information. And if you want to take best advantage of color in your photos, it helps to know a little about how digital color works. Before we can talk about digital color, however, we have to talk a little about color in general.

Mixing Colors: How Cameras and Monitors Work

One thing most of us learned early on in life, when finger painting was the height of our artistic achievement, is that when you mix two colors together, you get a third color. Somewhere along the way you may also have paid enough attention in an art or physics class to learn that if you start with the right colors, you can use just three colors to create all the other colors you can see.

There happen to be different combinations of colors that will work in different situations. For printers, the preferred starting colors, or *primary colors*, are cyan, yellow, and magenta. (Cyan is a light blue; magenta is a purplish red.) For digital cameras—as well as monitors, scanners, and film cameras, for that matter—the primary colors are red, green, and blue.

You may or may not be aware of how color televisions and computer monitors trick the eye into seeing color. (Both work the same way.) There are some variations on the theme, but, basically, the screens have sets of red, green, and blue phosphor dots, arranged as in Figure 2-8.

Digital cameras take pictures in RGB mode.

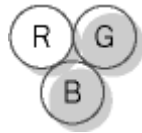


Figure 2-8.

Monitors create all their colors with just red, green, and blue phosphor dots.

RGB images have three channels: red, green, and blue. When the channels are shown separately, you get three black and white images, as shown in Figures 2-9 through 2-11.

Monitors and televisions can't mix colors the way you used to mix fingerpaints, because the dots are permanently in place. What monitors and TVs do instead is make the different color dots glow at different intensities to add more of one primary color and less of another. The dots are close enough together so that instead of seeing separate dots, the eye mixes the colors together to come up with the colors you see.

Now, pretend for a moment that color digital cameras don't exist and your boss has given you until next Tuesday to invent one. One thing you could do is link three monochrome cameras together so they can take the same picture at the same time, then put a red filter in front of the first camera, a blue filter in front of the second, and a green filter in front of the third.

When you take the picture, you'll have three different monochrome pictures, all in black and white. However, they'll look different because the first camera saw only the red in the scene as in Figure 2-9, the second camera saw only the green as in Figure 2-10, and the third camera saw only the blue as in Figure 2-11.



Figure 2-9.

In this red component of a color photo the barn is lighter than it is in the green and blue components because the barn is red.



Figure 2-10.

In this green component the grass is lighter than it is in the red and blue components because the grass is green.



Figure 2-11.

The blue component of the same photograph has nothing particularly notable in it.

Once you have these three images, all you need is some software that will show all three together at the same place on your screen, one effectively on top of the other, with each one lighting up just one set of phosphor colors—red, green, or blue. The first one needs to show as a monochrome red image (red phosphor dots only), the second as a monochrome green image (green phosphor dots only), and the third as a monochrome blue image (blue phosphor dots only). Since your eye will mix the colors together, you'll wind up with a full-color image. And that's basically how digital cameras (and monitors) work.

Color as Shades of Gray

Not so incidentally, the way we got the red, green, and blue components of the photo in Figures 2-9 through 2-11 was almost the reverse of the process we described to create a color image. We started with a full-color photo. Then, for each of the figures, we edited the original image in a photo editing program to eliminate the other two colors and show the monochrome image in black and white.

This ability to start with three black-and-white versions of a color scene and end up with a full-color image has led people who work with this stuff every day to talk about color in terms of *shades of gray*. So if someone says that a camera or monitor can produce, say, 256 shades of gray, what they mean is that it uses 256 shades of red, 256 shades of green, and 256 shades of blue. And, by the way, if you're interested in a black-and-white image, it can also give you 256 shades of actual gray.

This idea of referring to shades of colors as shades of gray is a useful shorthand, and it's one we'll use, too. However, there's more to it than just a shorthand description; there's a recognition that there is a certain equivalence between black-and-white images and color images. More precisely, if there are too few shades of gray to reproduce a color image well, the same flaws will show up in a black-and-white image with the same number of shades.

That's an important point: much of what's true for black and white is also true for color. And that means you can simplify discussions about color by talking about black-and-white images and shades of gray instead of having to talk about shades of colors and trying to visualize how the three primary colors interact with each other. More important for our purposes—since we're limited to using monochrome images in—it means the examples you'll see here in monochrome apply just as well to color images.

Digital Color

We've snuck it in here without any fanfare, but you may have noticed that suddenly we're back to talking about steps. When we talk about 256 shades of gray, what we're talking about is 256 individual steps: black, white, and 254 additional, discrete steps in between.

The human eye doesn't see the real world that way, of course. When you look at, say, a cloudless sky, the shade of blue changes between the horizon and straight overhead, but it doesn't change in sudden, discontinuous steps; it changes in a gradual, continuous way, much like the grays in the monochrome image in Figure 2-12 shade continuously from one to the next.



Figure 2-12.

This graphic shades from black to white, changing in a continuous way.

We're using a graphic image rather than a photo in Figure 2-12, because it lets us control the range of shades. In this case, the range, or *gradient*, goes all the way from black on the left to white on the right. However, the same sort of gradient shows up in all sorts of real-world objects and the photos of those objects. We already mentioned the blues in the sky. Flesh tones are another key area where shades change gradually and continuously. Figure 2-13 shows examples of both.



Figure 2-13.

In the real world, colors (reduced here to shades of gray) often change in a continuous way.

Obviously, digital photos don't (in general) show steps in this sort of image. If they did, you wouldn't expect many people to buy digital cameras.

The obvious question is, how small do the steps have to be so you won't be able to see them? The long answer is in the next section, "Shades, Colors, and Bits." The short answer is 256 shades of gray, even though you can sometimes get by with less. The gradient in Figure 2-12 uses 256 shades of gray, which means it should certainly not show any steps.

By the way, don't confuse the number of shades of gray with the number of colors. If you've ever used 256-color mode on your computer, you may have noticed that it's not suitable for viewing photos. The shading is nowhere near continuous, and you can easily see the steps as discrete colors with sharp boundaries going from one shade to another. That's because 256 colors is not the same thing as 256 shades of gray. It's actually close to 6 shades of gray. (The next section also explains the relationship between these two approaches to counting the number of colors.) When a photo shows obvious steps in the shading, it's called *posterized*.

Shades, Colors, and Bits

As we mentioned in the last section, the number of shades of gray is not the same thing as the number of colors. However both numbers will tell you the *color depth*, or how many colors are available in a given digital scheme, with more colors translating to a greater, or higher, color depth.

There's a straightforward relationship between these two ways of giving color depth, and it's useful to know how to translate from one way of counting colors to the other. A slight complication, however, is that there's a third way of counting colors too, and we need to cover that third way first.

If you've ever adjusted the resolution or other video settings on your computer, you may have changed the setting for the number of colors as well, or at least noticed that you can change the number if you want to. Some games, for example, require that you switch to 256-color mode, after which you may want to switch back to something else.

The choices you have available—and what they are called—depend on your video card, your monitor, your version of Microsoft Windows, and the resolution you're using. They also depend on where you look, since some choices are hidden away where you won't usually come across them.

Keeping that hedge in mind, a typical list includes 16 colors and 256 colors at the low end of the choices, and then switches to talking about the number of bits instead of the number of colors, most often with choices for 16 bits and either 24 bits or 32 bits or both. Adding still more room for confusion, the 16-bit choice is sometimes listed as high color or medium, and both the 24-bit and 32-bit choices are sometimes listed as true color or highest color.

The good news is that all these different approaches to counting colors—number of shades of gray, number of colors, and number of bits—are reasonably easy to sort out.

Bits and Colors

First, it helps to understand what a bit is. In addition to being digital, computers are binary, which means they represent everything with just two numbers: 0 and 1. Each of these numbers is a binary digit, which is usually abbreviated to *bit*.

Computers keep track of colors (or, more precisely, programmers tell computers how to keep track of colors) by assigning some arbitrary number of bits to create a code for each color. This works something like Morse code, which uses dots and dashes, and assigns an arbitrary pattern of dots and dashes to each letter of the alphabet. For example, in Morse code, the pattern • - (*Dot Dash*) is the code for the letter A. The pattern - • (*Dash Dot*) is the code for the letter N. Substitute 0s for dots and 1s for dashes, and Morse code looks an awful lot like binary code: 01 for *Dot Dash*, and 10 for *Dash Dot*.

A *bit* is like an atom, a pixel, or a grain of sand. It is the building block of a larger entity and is the basis for an entire unit. A bit is the smallest unit of data that a computer uses.

Although a computer can work with a single bit, bits are usually grouped into eight bits, called a *byte*.

A *megabyte* is about one million bytes. A *gigabyte* is about a billion bytes.

In the case of binary codes for colors, the number of colors you can define depends on the number of bits you've assigned to keep track of colors. If you assign just 1 bit to a color, you have only 2 codes to use—0 and 1—and only 2 colors you can define—usually black and white, with no grays.

If you assign 2 bits, you have 4 codes you can use, which is enough to keep track of 4 colors:

00
01
10
11

If you assign 3 bits, you have 8 possible codes to let you track 8 colors:

000 100
001 101
010 110
011 111

And so on.

As you might guess, you don't have to use all the codes available for any given number of bits, but you can if you want to. With each additional bit that you assign to keep track of colors, you double the number of colors you can keep track of. This doubling of bits adds up pretty quickly. At 8 bits, the number of colors jumps up to 256, at 16 bits it's more than 65,000, and at 24 bits it's more than 16 million. It happens to be a little clumsy to talk about 65,536 colors or 16,777,216 colors, however. That's why, as you get to higher numbers of colors, both people and programs shift from talking about the number of colors to talking about the number of bits.

That should rarely be necessary, however. Although printers and products that use color liquid crystal display (LCD) screens—like LCD monitors and Pocket PCs—sometimes offer unusual color depths, the overwhelming majority of products that refer to color depth in bits offer one or both of just two choices: 16 bits and 24 bits. If you see mention of more bits, you can treat it as 24 bits.

You'll occasionally see references to color depths that are greater than 24 bits (notably 32 bits, 36 bits, and 48 bits), but if you take a close look at these claims, you'll see that what they're really delivering is 24-bit color in almost every case. The additional bits are either used for something else or aren't being used at all.

Some video cards, for example, offer a 32-bit mode, either instead of or in addition to their 24-bit mode. In either case, however, they're using only 24 bits for color. For those video cards that offer both, you won't see any difference in color if you switch Windows to use one mode or the other for the simple reason that there won't be any difference. The 32-bit mode is still using 24-bit color.

Most games (and some third-party imaging programs) require you to have a specific type of video card that can handle specific color modes. Some require that those cards also have on-board RAM (random access memory).

Most of the other products that claim to use more than 24 bits are scanners. In some cases, the claim is pure marketing hype because the extra bits don't carry any useful information. Even for those scanners that can take advantage of the additional bits, however, the information is used internally only. What gets handed over to the computer still uses only 24 bits for color.

Why 24 Bits Is a Magic Number

There's a good reason why these color depths are the most common choices.

First, the more bits that your camera, computer, printer, or video card has to shuffle around, the more work it has to do and the longer it takes. More bits also means taking up more space in your camera and on your hard disk. So unless you like waiting for your camera and computer to catch up to you, and you like running out of room, you don't want to use more bits in an image than you need.

The question then becomes what is the smallest number of bits you can get away with. Ask a color scientist, and he or she will tell you that people can see something in the very rough range of 8 million distinct colors. A close fit between the number of colors you can see and the number of codes available is 23 bits.

Unfortunately, assigning the 8 million or so codes to just the right colors would not only be extremely time-consuming, it would be tricky beyond belief. You'd have to assign the colors carefully, because the human eye is more sensitive to changes in some colors than in others, and because the number of colors you can see varies from one person to the next.

Jumping up to the next level, at 24 bits, solves that problem. By tracking twice as many colors as you actually need, you don't have to worry about which colors the eye is more sensitive to or whether some people can make fine distinctions that others can't see. Just assign the codes so the colors are spread out evenly, and you'll automatically cover all the bases. This doesn't mean that you can't wind up with poor color using 24-bits, just that it's hard to mess up.

In general, color in photos looks good enough at this level for 24-bit color to have earned the name *true color*, which you can define as being a match for film photography, or, in context of this discussion, as the level of digital color that won't show any steps. And that's why 24 bits is one of the most common choices for color depth.

Alas, not all computers—particularly older computers—are fast enough to run at tolerable speeds using 24-bit color. The solution is to pick another level of color depth that speeds things up significantly and still gives pretty good color.

As it happens, 16-bit color fills the bill. It leaves your system with only two-thirds as many bits to move around, and offers a level of color called *high color*, which means it has enough colors so a fairly large percentage of photos won't show any steps where they shouldn't.

Monitor color depth depends on the capacity of your hardware, including your monitor, your video card, and the software drivers that your card uses.

File color depth depends on the file format you choose for taking and saving your digital pictures.

Even if your monitor does not support 32-bit color, you can still save in that format if you have the proper software.

Now that we've covered some of the digital technology basics, let's take a look at digital cameras.

Quick Quiz

1. What is the essential difference between analog and digital technologies?
2. When drawing a line on a piece of paper using a pencil, are you drawing something that is analog or digital?
3. True or False? When digital technology can break down its "steps" far enough, the differences between analog and digital are effectively erased, and the human eye cannot discern the difference between the two.
4. What is the easiest way to view the actual pixels in a digital image on a computer monitor?
5. What is a pixel? What is a megapixel?
6. What causes pixelization in a print of a digital image?
7. How can you avoid pixelization when printing digital photos?
8. What three colors do monitors use to display images?
9. What real-world examples are there of gradients?
10. Name three things that determine what level of color depth you can get from your monitor.

Putting It All Together

Exercise 1: A circle is a continuous shape. Continuity is generally associated with analog technologies. However, when drawn at the computer, the circle is not an analog piece of data; it is digital. Using Microsoft Paint, draw a circle and show its digital attributes.

Exercise 2: Explain in your own words why 24 bits is the magic number for monitor displays, Pocket PCs, video cards, scanners, and similar hardware, and then change your monitor's display (or verify it is set to or cannot be set to) 24 bit display.