TIFF, GIF, and PNG: Get the Picture?

Jeff Kabachinski

What's the difference between graphic and picture file formats? You've seen them—TIFF, GIF, PNG, JPEG, and BMP are among the most popular formats. In this issue, IT World surveys these file formats by looking into the history, how and where they're used, and what each one is most suited for. The story of digital picture formats is one of compression, true color, and protest campaigns. Read on!

TIFF

Tag Image File Format (TIFF) was developed in the mid-80s by the Aldus Corporation. The original intent was to be as flexible as possible when exchanging image data between machines and applications. The current specification for TIFF, created in 1992, is controlled by Adobe (version 6.0) and can be found at http://partners.adobe.com/public/developer/tiff/index.html. TIFF is popular for high color depth pictures and line art. TIFF is also supported by all the major graphics and image manipulation and desktop publishing applications as well as word processing and scanning programs.

TIFF started out as a common format for desktop scanner companies. Originally a simple bi-level image format (black and white), since that was all the scanners of the day could handle, TIFF evolved first into grayscale and then into color.

TIFF is flexible and adaptable. A single file can contain multiple images and data via tags in the file header. Tags include basic image sizing information as well as any compression means that were implemented. A TIFF file can also contain JPEG compressed images. One of the best things about TIFF is that it contains all the original data, even if edited. This is known as lossless.

TIFF also can use lossless data compression techniques like LZW. LZW comes from the authors' names—Lempel-Ziv-Welch. Actually it is Terry Welch's improved implementation of Abraham Lempel's and Jacob Ziv's LZ78 program, first launched in 1978. It was meant to be quickly implemented and therefore suffers a little bit from not being overly thorough. The LZW process, also used by GIF files, builds an ordered sequence of symbols or codes (called a string) from the data being compressed. It then takes two characters from the string and enters them as a concatenation relative to the first character of the short string. In effect we can reduce the number of bits in the file by encoding them in character strings and saving the strings. LZW is known as lossless compression. We're not throwing any information away—we're just encoding it in pairs and have a scheme to remember what's in the pairs to make the sent file smaller.

Some compression techniques are "lossy" in that some data are changed and/or removed. If we removed the two least significant bits in each byte of the data, the picture may still look

Check Points

What format is best for your photos and images? It depends…

✓ A TIFF is best for archiving pictures; just save a copy before you modify it so you always have a master.

✓ JPEG, GIF, and PNG files are better for e-mailing and websites because of their small size—a trade off for good quality.

good enough, but we removed 25% of the data and file size. And with lossy techniques like this, we can't go back to the original data stream. It's lost forever.

Each TIFF file begins with a 2 byte code that indicates if the data within are in big or little endian mode. This indicates the byte order of hexadecimal information. It comes from the Apple/MacIntosh and Windows based computer operating systems. Contrary to popular belief (at least in some circles), the ultimate answer is not 42. However, using 42 in base 16 (66 in decimal) as our example, 42 is the big-endian mode expression (MSB first) and 24 is the little-endian mode expression (LSB first). We usually express numbers in big-endian mode in English—forty-two rather that two and forty. To go further off track, the term big-endian comes from Jonathan Swift's novel *Gulliver's Travels*. In the story, tension arose from the Big-endians as they preferred to crack open their

soft-boiled eggs from the big end, contrary to royal edict. Hence they became known as Big-endians.

But back to TIFF. The last thing to mention about TIFF is that due to its lossless aspect, it is a great method for archiving pictures. And if you save your original in that format, you'll always have a "golden copy."

GIF

Graphical Interchange Format (GIF), was first introduced by CompuServe in 1987. It's become popular due to its ease of portability, especially online. GIF is an 8 bit per pixel bitmapped image format. This is also referred to as raster graphics, because the rows of pixel informa-

Dots, Lines, Pixels, and Samples

- DPI—Dots per Inch often refers to the printing resolution. Increasing an image's DPI will depict more shades and picture details, but the image's dimensions remain the same. A 600 DPI laser printer can print up to 600 dots of picture information per inch. However, a 600 DPI photograph will look the same when printed by a 300 DPI laser printer as a 300 DPI photo would. The 600 DPI photo, however, will have a larger file size.
- LPI—Lines per Inch also refers to the way printers produce the image. For a picture to look more even or for softer fades, the number of lines or rows needs to increase. You'll see resolutions at 50 LPI for laser printers, 133 LPI for uncoated stock, 150 LPI for coated stock, and in the 175 to 200 LPI range for color in magazines. This is why you need special coated stock for your picture printer.
- PPI—Pixels per Inch refers to the resolution of the image as seen on the screen or monitor. Digital cameras create images at a fixed number of pixels. A digital camera with an image resolution of 1200 x 1600 pixels will provide a "photo-quality" printed image at 4″ x 5″ but a lower quality print for an 8″ x 10″.
- SPI—Samples per Inch refers to scanner and image resolution. The higher the SPI, the higher the resolution.

tion represent how they appear on your monitor screen, which uses a raster to light up the pixels. This format uses 256 colors in its Red-Green-Blue (RGB) color space. It also supports animation and can specify a different color palette for every frame. It is best suited for things like clip art images and logos with large areas of solid color. However, due to the ability to store a new palette each time, it can be used fairly successfully for pictures as well.

The two main versions of GIF are 87a (from 1987) and 89b (from 1989, when CompuServe decided to enhance GIF capabilities). You can tell which version is used from the first 6 bytes of the file, which, in American Standard Code for Information Interchange (ASCII) spells out either "GIF87a" or "GIF89b." GIF89b added the ability to store multiple images in one file, which makes small animations possible. If you've seen the waving flag that some folks have in their e-mail signature—that's an animated GIF.

A GIF is also useful for sharp-edged line art due to its compressibility and smallish color palette. While a new palette can be used for each frame within a file, the maximum number of colors remains at 256. This was perfectly reasonable at the time (1989—ancient history as far as graphics go) as higher color was not readily available. In addition, dithering was possible in the frame. Dithering means simply patterning color dots to approximate a new color. For example, dithering, or combining and patterning black and white dots can approximate greyscale.

The biggest advantage to GIF is its relatively small file size. This means the image is built on the computer screen quickly and saves memory space. It means transmitting it across the internet quickly as well. For these reasons, GIF has become the most popular method for displaying images on websites. GIF images can also be interlaced to make images appear even faster. You might recall back in the days of dial-up internet access, going to a website and seeing the picture build from a rough, crude looking image "melting" to a higher resolution. That was the interlacing effect of building the image on the screen in stripes.

GIF is considered lossy. Essentially that means when a GIF is compressed, the original image information is not retained. In other words, if you make a change to the image, you're stuck with the change unless you've kept a copy of the original. That's the main tradeoff for smaller file size.

GIF can also use the LZW data compression utility. Originally, LZW was patented by Unisys, but Com-

puServe built GIF to use LZW, unaware of the patent. Unisys wanted to license the use of LZW to all major information services companies, including those with on-line service (it did not, however, expect noncommercial, nonprofit companies to pay). This in effect led to another file format called PNG, developed in the mid-90s as a replacement to GIF.

The PNG Alternative

Portable Network Graphics (PNG) overcame some of GIF's deficiencies. The main aspect PNG brought is scalability of color depth, as it allows 16, 24, and 32-bit images. A dual edged sword, this also means larger file size and longer download times. PNG has more features and a higher compression utility, but it does not offer the animation aspect that GIF has. Probably the coolest thing about PNG is "alpha transparency." When PNG is used in its 32-bit format, it uses 8 bits for red information, 8 bits for green information, 8 bits for blue information, and 8 bits as an "alpha-channel." The alpha channel is used for transparency information. When you overlay pictures, the one on top will allow the picture below to bleed through. The amount that bleeds through is based on what the transparency level is.

While PNG continues to gain favor for web use, it hasn't really taken off, due mainly to the animation that GIF can provide. In addition, the $5,000 to $7,500 license fees for GIF expired with the patent in 2003 in the United States and 2004 in other countries, so GIF can now be used freely. An interesting side note, the League for Programming Freedom (lpf.ai.mit.edu) had a "Burn all GIFs" campaign in protest of Unisys's selective enforcement of GIF licensing fees.

JPEG

The Joint Photographic Experts Group (JPEG) is the name of the committee that started the JPEG standard. Launched in 1986, it didn't publish the first JPEG standard until 1992. It defines both the compression-decompression (codec) and the file format for photographs. Codec defines how the photo is compressed into a stream of bytes as well as how to decompress the image to present it. The big issue with JPEG is its lossy compression—lossy to the point that there are warnings about "JPEG rot." Say you have a photo of your family, but the baby's eyes have that devilish red spot in them. So you open your photo fixing software and remove the red eye. It asks you, are you sure? If you click "Yes," now the file is saved with these changes. At this point, you can't go back to the original. Do some cropping, add a border,

	PNG	JPEG	GIF
Full Name	Portable Network Graphics	Joint Photographic Experts Group	Graphical Interchange Format
File Extension	.png	.jpg, .jpe, .jpeg	.gif
Colors	16,777,215	16,777,215	256
Transparency	256 levels of transparency per pixel	None	One color may be transparent (optionally)
Compression	Lossless—up to ~7:1	Lossy—up to 20:1	Lossless—starting at 3:1 with LZW can get to 10:1
Differentiator	Gamma correction as part of the PNG spec (see below)	Progressive rendering	Interlaced and/or animated
Workings	Compared to a linearly encoded image, PNG allocates fewer sample values to brighter parts of the tonal range and more sample values to the darker portions of the tonal range to compensate for a cathode ray tube's (CRT) projection of the image.	Designed to compress or reduce things that don't make a lot of difference in what you see. Based on human studies, it throws away things we can't see anyway.	It uses a color palette values instead of putting 24-bit values into a map.

Table 1. An overview of picture file formats.

change the contrast or brightness—all changes permanently change the file. In addition, every time you save the file, it gets re-compressed and degrades a little bit. Every time you open and close the file, JPEG rot occurs. The solution? Save it in a different format. The experts say that PNG is best. It's quite a bit smaller that TIFF or Windows bitmap (we'll see a bit later). The tip from the experts is to save a master copy in full color PNG unless you have one of the better graphics applications like PhotoShop Pro that has its own lossless file format (*.psd, for example).

JPEG's codec is an important feature when used online or over a network connection. JPEG is said to be symmetrical in that it needs equal time on the transmitting side (compression) as on the receiving side (decompression). It works by converting a color image into rows of pixels. Each pixel is defined by color and brightness levels. The picture is broken down into 16 x 16 pixel blocks, and then compressed into an 8 x 8 bit pixel block by subtracting every other pixel. It further compresses the data by using an algorithm that determines an average value for the block. The data can be further compressed—all in the spirit of smaller amounts of data that can be transmitted faster. Eventually compression levels of 20:1 or 30:1 can be achieved. To decompress, a reverse procedure is applied. Picture quality can suffer to the point of "tiling" where the picture appears to be a mosaic of tiles due to the pixel blocks that JPEG uses.

Other versions of JPEG (such as JPEG 2000 and JPEG++) are becoming popular. In the case of JPEG 2000, even lower levels of resolution can be achieved. For example if you're surfing with a handheld wireless device, you can't really see the picture detail anyway, so why take the time to try and show it? JPEG 2000 simultaneously allows for higher resolution if you've got a fat connection and a 21-inch hi-res monitor.

The details of the compression techniques that are part of JPEG are beyond the scope of this issue of IT World. You can easily find all the grim details with minimal google searches. You can find out about block splitting, downsampling, quantization, and entropy coding to delight your friends and neighbors with this information. Maybe that's why I don't have too many friends and neighbors stopping by much anymore.

BMP

Microsoft Windows bitmaps (BMP) is another raster graphic format that saves the image as a rectangular grid of pixels.

BMPs first appeared in the MS Windows platform but are now widely recognized. BMP is a Device Independent Bitmap (DIB), where the format for color information is independent of the output method or device. Therefore, there must be a device driver to translate the image information for display. BMP is also a 24-bit color and grayscale system for portraying graphical images, including pictures. BMP in its native format is not compressed. Therefore, a 1024*768 image would be nearly 2.4 megabytes.

Run Length Encoding (RLE) compression techniques can be used with BMP and most raster graphics formats. A pretty simple technique, RLE encodes redundant pixel string information by indicating the run length. For example, the pixel string or run of HNCCCCBDDDDEFFFFF would be represented as HN*4CB*3DE*5F, where the * symbol indicates the following information is RLE compressed. The compression of our example is taking 16 characters down to 13, not too bad for such a short string. The compression ratio depends on how redundant the information is. In a fax, where most of the BMPed image is white space and can be relayed as long runs of white pixels, the compression ratio is quite high.

Summary

GIF, JPEG, and PNG are most likely the best formats to use for three reasons. First, they're standardized and open formats for anyone to use. In addition, JPEG is an ISO standard and PNG is an IETF RFC (Internet Engineering Task Force Request for Comments—www.ietf.org) and W3C recommendation (World Wide Web Consortium—www.w3.org). Second, they're compressible. GIF files are generally compressed at 5:1, JPEG at 10:1 or 20:1 and PNG at about 7:1. Finally, they're all supported by web browsers. Well, pretty much. Microsoft's Internet Explorer doesn't support the alpha channel transparency for PNG—but, on the other hand, GIF and JPEG don't have the alpha channel at all. Use TIFF to archive your original pictures as it is a lossless format. Check out the summary table and sidebar for more information regarding these picture file formats. ∎

Jeff Kabachinski, MCNE, MS-T, is manager of workplace learning and performance for GE Healthcare located in Waukesha, WI. He holds MCNE certification, an MSc in training and development from Leicester University, and a BS in engineering technology from the Milwaukee School of Engineering. E-mail: Jeff.Kabachinski@ge.com.