Once you start scanning images and reproducing them as halftones, you find out very quickly that the file sizes can be enormous. It is therefore important to keep the file size as small as possible without hurting the final quality. The purpose of this article is to present some background information about the scanning of images, and also to present some guidelines on how to keep those scanned images reasonable in size.

Raster and vector images

Scanned images are often referred to as raster images. Images created in drawing programs like Adobe Illustrator** are called vector images. The distinction between these two types of images is very important. Raster images are created either in a paint program or by scanning. Either way, they are described mathematically as points on a grid. If you enlarge this grid, you may start to see the jagged edges that are characteristic of a raster. Vector images, on the other hand, are files that have as their basis mathematical formulas describing lines and curves. They may be enlarged or reduced without affecting their output quality.

Raster files are larger in size than vector files because each piece of information on the grid has to be stored. The finer the grid, (i.e., the higher the resolution,) the more points on the grid, and therefore the larger the file. Vector files tend to be smaller because the formulas used to create them may be stored compactly. (For example, you can describe a line simply by naming the two endpoints.) Their small size however, does not necessarily mean that they will output faster; their complexity is of a different type (but that is another story.) As a summary:

	Vector	**Raster**
Also called	Draw, object-oriented, resolution independent	Paint, bitmapped, scanned, resolution dependent
Sizing	Sizing doesn't affect quality	Sizing effectively increases or decreases the image resolution
File Size	Relatively small (usually under 300 KB)	Potentially large (from 1 to 50 MB or more)
Categories	Line art, synthetic art	Paint images, scanned line art and halftones
Formats	EPS[1]	Pict, Pict 2, TIFF, RIFF

[1]EPS files may sometimes contain raster information. This is the case with some color-separated EPS5 files which contain a main (or view) file, and files for the cyan, magenta, yellow, and black separations.

Although raster and vector files are inherently different, recently several programs have been released that are capable of converting from raster to vector or vice versa. Adobe Streamline** can turn a raster image into vector line work that may then be reworked in Adobe Illustrator. One feature of ColorStudio** is the ability to turn vector line work into a rasterized image.

Bit, byte, kilobyte, megabyte...?

This article deals solely with the file sizes of raster, or scanned, images. File size in scanned images is measured in bits, bytes, kilobytes, megabytes and gigabytes. A bit is the smallest piece of information that can be stored in a

digital computer. It can distinguish between two things: yes or no, black or white, on or off, etc. The next category is the byte which contains eight bits. There are 1024 bytes in a Kilobyte (or KB). There are 1,048,576 bytes in a Megabyte (or MB). There are 1,073,741,824 bytes in a Gigabyte (or GB). Each of these steps increments by a factor of 1024.

Size	Could contain	Fits on
1 bit	One state (Yes or No)	1 memory location
1 byte	One character (a letter or a number)	8 memory locations
1 kilobyte	A paragraph	Diskette[2]
1 megabyte	A 3 x 4", 300 dot per inch, scanned B&W image	Internal hard disk[3]
1 gigabyte	12, 8 x 10", 600 dot per inch, scanned color images	Two optical disks[4]

[2]Most diskettes hold around 800 KB, double density diskettes may hold up to 1.4 MB

[3]Internal hard disks can hold 20 to 80 MB, removable hard disks hold around 40 MB

[4]Optical disks hold 300 MB on a side, external hard disks commonly hold 300 MB

Resolution

The resolution of a scanner is the key determining factor in file size. To truly understand scanner resolution you have to think of it as if it were three-dimensional. For imagesetters, resolution is only two-dimensional. It is dependent on the number of dots in the horizontal and vertical directions. Scanners have not only the horizontal and vertical directions, but also a third dimension, or depth. Both scanners and imagesetters describe horizontal and vertical resolution in terms of dots per inch (sometimes pixels or samples per inch), but depth is only found in scanners. It describes the number of bits of color or grayscale information per pixel. The horizontal and vertical resolutions are often referred to as spatial resolution, while depth (of grayscale or color) is referred to as tonal resolution. When the term "resolution" is used alone, it usually refers to spatial resolution.

A line art scan is only one bit deep because one bit is all that is needed to distinguish black from white. Grayscale and color information require more bits. Most grayscale images are 8 bits deep (that's enough to distinguish 256 levels of gray). Color images are commonly 24 bits deep (enough to distinguish 16.7 million different colors).
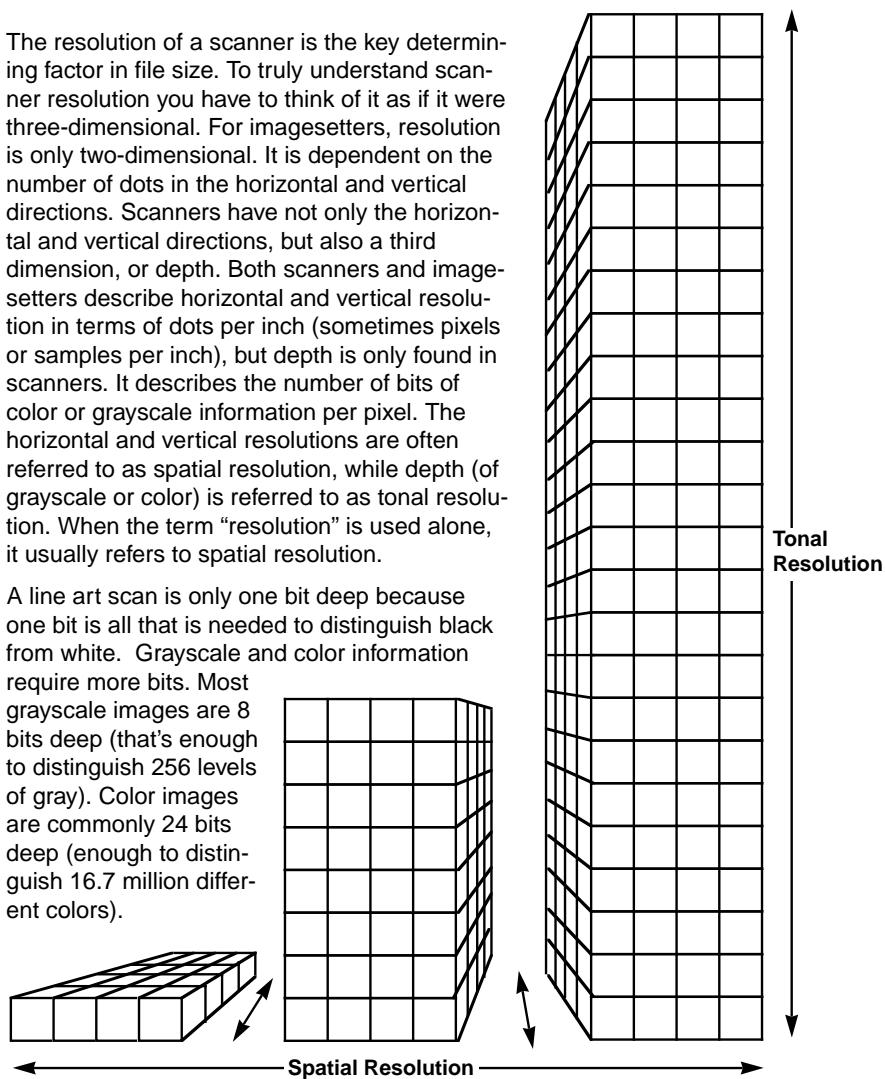

Figure 1 - Resolution shown three-dimensionally.
The set of blocks on the left represent an image with one bit of tonal resolution, the center image is a grayscale image with 8 bits of depth, and the one on the right represents a 24 bit color image.

Line art and halftones

The tonal resolution setting of the scanner is what makes the difference between a line art scan (1 bit) and a B&W halftone scan (usually 8 bit). But the approach to scanning line art and halftones should also be different. For line art, the spatial resolution ideally should match the resolution of the output device. So to output raster line art on an imagesetter at a 1270 dot per inch resolution, you might want to scan at 1270.[5] For a halftone you don't need to scan at such a high spatial resolution. At a minimum, you need one piece of information for each halftone dot. If your screen ruling is 150 lines per inch, you could scan at 150 dots per inch. But there is more to it than this.

[5]One of the real drawbacks of scanning technology is in high-resolution line art scanning. There simply aren't any low-cost, high-resolution line art scanners. The best option is to create line work with a vector drawing program.

Sizing and resolution
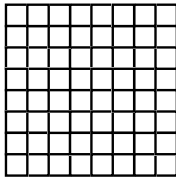
One square inch of a
8 dot per inch scan
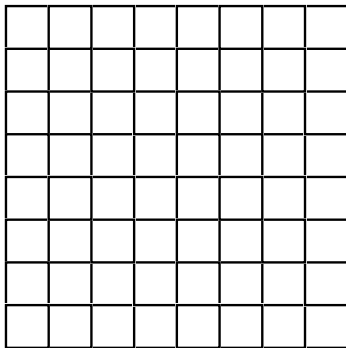
The same scan
enlarged two times

Figure 2 - Sizing and resolution

Because raster images are resolution-dependent, if you re-size an image, you effectively change the spatial resolution. It works like this: say you scan at 8 dots per inch and then blow that image up twice in size so that what was one inch is now two inches. If you had 8 samples[6] in an inch before sizing, you now have 8 samples in two inches after sizing. 8 samples in two inches means that your spatial resolution is now 4 dots per inch. (See Figure 2.) Re-sizing affects only spatial resolution, not tonal resolution.

[6]The terms dots, pixels, and samples are often used interchangeably. Dots per inch is used to describe resolution in both imagesetters and scanners. Pixels (or samples) per inch, and bits per pixel (or sample) are used to describe spatial and tonal resolution respectively. The term sample also gives a more accurate description of what the scanner is really doing: looking at tiny areas in the image and taking a sample.

The following formula allows you to calculate the effective spatial resolution based on the initial spatial resolution and the amount it has been sized:

Initial spatial resolution / Sizing factor = Effective spatial resolution

For example: 8 dots per inch / 2 = 4 dots per inch

To calculate the sizing factor, take the new size and divide it by the original size, for example: 2" (the new size)/1" (the original size) = 2 (sizing factor). In the case of enlargements the spatial resolution decreases; reduction, however, actually increases it! Here are some common sizing factors:

Sizing factor	Percent change	Called
4	400%	Four time enlargement
2	200%	Two time enlargement
1.5	150%	1 1/2 time enlargement
1	100%	Same size
.75	75%	3/4 reduction
.5	50%	1/2 reduction
.25	25%	1/4 reduction
.1	10%	1/10 reduction

Choosing a resolution

Good planning is critical in the scanning of a grayscale image. Before you scan, you should know the halftone screen ruling that the image will be reproduced at and the size that it will be enlarged or reduced to. If you don't know either of these specifications, you're in a difficult position. You can always scan at the highest spatial resolution possible, but this will make the file size extremely large, and you may regret this later during output. A more sensible approach is to use a low-resolution scan for sizing and placement until the screen ruling and sizing are known. Then the scan can be made at the proper resolution. For example, say you know that you want to scan an image that will be printed at a 150 line per inch screen ruling, and enlarged two times (from a 4 x 5" original to a 8 x 10" reproduction). Here's how you would go about choosing the scanner spatial resolution. If you want an effective spatial

resolution of twice the screen ruling (2 x 150 = 300), you can use the formula to work backwards:

(Initial spatial resolution)/2 = 300 dots per inch
Initial spatial resolution = 2 x 300 = 600

You should then set your spatial resolution on the scanner to be as close to 600 as possible. In some cases, the number you calculate here may be higher than the highest spatial resolution setting on your scanner. If that is the case, you have reached an enlargement size that is too large for your scanner. There are limits as to how much you can enlarge images, and that is solely a function of the highest spatial resolution of the scanner, the screen ruling of the image, and the ratio of scanned samples to halftone dots.

The scanning ratio

Before you scan and print an image, you will have to make quality decisions. First, how many samples do you want for each halftone dot? This is called the scanning ratio, and it is based on the spatial resolution setting of the scanner. You want to be sure that you have at least one sample for each halftone dot. Below a 1 to 1 ratio, you risk the chance of an artifact called pixelization. (See

Figure 3 - Pixelization

Figure 3.) According to sampling theory, above a ratio of 2 to 1 you don't get any significant benefit from the extra data. So if you don't benefit from it, what is the sense of scanning it, storing it, and moving it around a network?

Output time

Common sense tells you that a large image will take longer to output than a small image. But how much longer? It would be convenient if there were a formula which showed how long a scanned image would take to output based on its file size. Unfortunately there isn't. So many factors are involved (RIP, imagesetter, network) that it is hard to judge. However, we are able to give you some figures for a specific image and a given set of equipment. A sample accompanies this technical information piece, which includes examples of output time, file size and scanning ratio. If you didn't receive this sample, please contact me at the number listed at the end of this document.

One easy way to improve output speed is to crop your images as close as possible to the final size before you bring them into your page layout program. If you don't do this the RIP has to process the entire image, rather than just the portion you want printed, and this slows output.

Compression

Because file size can become unmanageable, file compression is an important issue. As it stands now, some programs allow you to compress images when you store them. These compression schemes may be loss-less (i.e., no image degradation) or loss-y (some image degradation, but much greater compression.) PostScript** will provide even more effective compression in the near future. The compression scheme in PostScript Level 2 is lossy, but it also allows you to control the level of compression. Generally, as the amount of compression increases, so does the amount of loss. Users will be able to select a level of compression that meets their needs.

File Size Charts

It helps to be able to estimate the size of a file. You might, for example, want to keep the file small enough so it will fit on a diskette. Or maybe you are doing numerous images and want to see that they will fit on one optical disk. The charts on the following page show information on file size for some common image sizes and resolutions:

1 bit images - Line art

Scanner	Image Dimensions				
Resolution	**1 x 1"**	**2 x 2.5"**	**4 x 5"**	**5 x 7"**	**8 x 10"**
75 dpi	703 B	3.4 KB	13.7 KB	24.0 KB	54.9.KB
100 dpi	1.2 KB	6.1 KB	24.4 KB	42.7 KB	97.7.KB
150 dpi	2.8 KB	13.7 KB	54.9 KB	96.1 KB	219.7.KB
200 dpi	4.9 KB	24.4 KB	97.7 KB	170.9 KB	390.6.KB
300 dpi	11.0 KB	54.9 KB	219.7 KB	384.5 KB	878.9.KB
400 dpi	19.5 KB	97.7 KB	390.6 KB	683.6 KB	1.5 MB
600 dpi	44.0 KB	219.7 KB	878.9 KB	1.5 MB	3.4.MB

Note: Use the one inch column to calculate values for image sizes that are not listed (for example, if your image is 3 x 3" multiply the value in the 1 x 1" column by 9 to get the estimated file size.) If you resize the image, the file size (in KB or MB) will remain constant, but the effective spatial resolution will change. Remember, initial spatial resolution = effective spatial resolution until you resize! Once you resize, initial spatial resolution / sizing factor = effective spatial resolution.

8 bit images - Grayscale

Scanner	Image Dimensions				
Resolution	**1 x 1"**	**2 x 2.5"**	**4 x 5"**	**5 x 7"**	**8 x 10"**
75 dpi	5.5 KB	27.5 KB	109.9 KB	192.3 KB	439.5 KB
100 dpi	9.8 KB	48.8 KB	195.3 KB	341.8 KB	781.3 KB
150 dpi	22.0 KB	109.9 KB	439.5 KB	769.0 KB	1.7 MB
200 dpi	39.1 KB	195.3 KB	781.3 KB	1.3 MB	3.1 MB
300 dpi	87.9 KB	439.5 KB	1.7 MB	3.00 MB	6.9 MB
400 dpi	156.3 KB	781.3 KB	3.1 MB	5.3 MB	12.2 MB
600 dpi	351.6 KB	1.7 MB	6.9 MB	12.0 MB	27.5 MB

24 bit images - Color

Scanner	Image Dimensions				
Resolution	**1 x 1"**	**2 x 2.5"**	**4 x 5"**	**5 x 7"**	**8 x 10"**
75 dpi	16.5 KB	82.4 KB	329.6 KB	576.8 KB	1.3 MB
100 dpi	29.3 KB	146.5 KB	585.9 KB	1.0 MB	2.3 MB
150 dpi	65.9 KB	329.6 KB	1.3 MB	2.3 MB	5.2 MB
200 dpi	117.2 KB	585.6 KB	2.3 MB	4.0 MB	9.2 MB
300 dpi	263.7 KB	1.3 MB	5.2 MB	9.0 MB	20.6 MB
400 dpi	468.8 KB	2.3 MB	9.2 MB	16.0 MB	36.6 MB
600 dpi	1.0 MB	5.2 MB	20.6 MB	36.1 MB	82.4 MB

Note: If you are scanning 24 bit color images at 600 dpi, the total size of the image in square inches before resizing will be equal to the file size in megabytes. The file size for grayscale 600 dpi images will be 1/3rd the total size of the image in square inches. When all other factors remain constant, file size increases by the square of the spatial resolution (for example, increasing the spatial resolution by four, enlarges the file size by sixteen.)

You may also use these charts to estimate the effective spatial resolution of a given scan. Determine the amount of data per square inch by taking the file size and dividing it by the dimensions in square inches. Then look for the closest match to this value in the 1 x 1" column of the appropriate chart.

Slide scanner resolutions

Slide scanner arrays are commonly made up of a fixed number of elements (also called scanning sites or light-sensitive diodes). A common slide scanner array has 4096 by 6144 elements over the size of a slide (approximately an inch by an inch and a half). If you were to calculate spatial resolution by this number, you would think that slide scanners have very high resolution, over 4000 dots per inch.These resolutions seem very high until you start enlarging slides to reasonable sizes. If a slide is enlarged to 4 x 6" the effective spatial resolution drops to close to 1000 dots per inch, at 8 x 12" it would be more like 500 dots per inch. This may be more data than you really need. In fact,

some scanners allow you to scan for less than the full spatial resolution. The following chart shows the file sizes for these conditions: slide scanner with a 4096 x 6144 pixel array, tonal and spatial resolutions as noted, maximum size for a given resolution as noted.

Slide scanner chart

	100 dpi	**300 dpi**	**500 dpi**	**1000 dpi**
1 bit	29.3 KB	263.7 KB	732.4 KB	2.9 MB
8 bit	234.4 KB	2.1 MB	5.7 MB	22.9 MB
24 bit	703.1 KB	6.2 MB	17.2 MB	68.7 MB
Max. size	40 x 60"	13.33 x 20"	8 x12"	4 x 6"

Formulas

In general, charts are only useful if they use exactly the set of variables that you need. But in computing file size there are too many variables. Here is a formula that allows you to estimate the file size by using your own variables:

(Tonal resolution/8) x (Spatial Resolution)^2 x (Dimensions) = File Size

    (bytes/sample)   (samples/in^2)   (in^2)   (bytes)

If you take the tonal resolution (in bits) and divide it by 8 you get the number of bytes per sample; squaring the spatial resolution gives the number of samples per square inch; if you multiply that by the number of bytes per sample, you get the number of bytes in a square inch; multiply that by the size of the job in square inches, and you get the total number of bytes in a job. Divide that by 1,024 to get kilobytes or by 1,048,576 to get megabytes. For example, for an 8 bit, 300 dpi, 8 x 10" image:

$$((8/8) \times (300)^2 \times 80) / 1,048,576 = 6.9 \text{ MB}$$

Working backwards you can use this same formula to find out the effective spatial resolution of an image by knowing its tonal resolution and file size (as long as the file hasn't been compressed):

Effective spatial resolution =
The square root of ((file size in KB x 1024)/(dimensions x tonal res./8))

Note: If the file size is in MB, multiply by 1,048,576.

You might want to do this if you have resized a file, and no longer know if the scanning ratio will be adequate.

Conclusion

A knowledge of the concepts of file size forms a basis of understanding that allows you to keep file size as low as possible, while maintaining the desired level of quality.

Comments

Please direct any questions or comments to:

Jim Hamilton, Marketing Department
Linotype-Hell Company
425 Oser Avenue
Hauppauge, NY 11788   (516) 434-2717