Image Basics

Shopping for Scanners

In your hunt for a scanner, the very first thing you need to do is figure out which features you need, and how much you want to pay.

There are lots of scanner review sites, so you might start there. If you're a PC user, there is a huge selection of scanners available. In the Mac realm, there are plenty of choices, too.

I bought my scanner, a HP ScanJet 3300C, for about US$100 last year. It has 36-bit color (I'll explain that later) and 600 dpi by 600 dpi resolution. In typical tech industry fashion, this scanner is already an antique, and is going for spare change on Ebay. But it still suits my simple needs, makes good reproductions, and is extremely easy to setup and use, which counts big in my book.

When you start looking at scanners and comparing features, you'll notice that the mumbo-jumbo factor is high — that is, there's lots of jargon to absorb. But don't be daunted. The scanner manufacturers lingo is easy to learn.

Scanner Jargon

When it comes to scanners, the two crucial measurements are bit, or color depth, and resolution. Both of these measures have to do with reproducing colors and grayscale in images, but refer to different aspects of the process.

Bit Depth

"Bit depth' refers to the number of bits of image data the scanner will use for each pixel. To create "true color" (more jargon), a scanner needs a 24 bit depth, which can yield 16.7 million colors (24 bits = 256 by 256 by 256 red, green, and blue dot combos). This is more than the eye can see, and most computer monitors are set to 24-bit color, so that's the max that can be displayed from most machines. These days, it's hard to find a scanner that doesn't have at least 24-bit capability.

However, if you want to spend the money, you can get scanners with a bit depth of 30, 36, 48 and more. With more depth, the scanner is able to sample more colors, which gives you more flexibility in adjusting the brightness and contrast. For instance, a 30-bit scanner has up to 1 billion potential colors.

Resolution/DPI

Resolution is measured in dots per inch, which tells you how many spots of information there are in a given area. So the DPI of a scanner tells you how fine a grid of pixels will be used to represent your images. The more dots per inch a scanner employs, the more detail the resulting image will have.

Before you scan an image, you will set the scanner for the appropriate DPI setting, which depends on how you plan to use the file. If you want to scan a sketch or

something similarly simple, the scanner only needs to sample a few colors and a low dpi will suffice. However, if you want to scan a colorful print with lots of detail, then it will take more pixel samples, a higher DPI, to accurately represent the image.

As a baseline, consider that the typical computer, and Web pages, display images at 72 or 80 DPI. Photographs in magazines are reproduced at about 200 to 300 DPI. Most scanners exceed 1200 dpi, so they can handle just about anything you throw at them. A typical rating for a mid-range, mid-priced scanner right now is about 1,200 by 2,400 DPI.

If you're going to send your file to someone over e-mail, it is best to scan at 150 DPI, or even less. That should give you decent resolution with a reasonable file size. If you want to scan something for print output, a higher DPI is appropriate, anywhere from 200 to 300 for starters. If you plan on resizing an image, a higher DPI will give you more detail and clarity when you blow it up.

If you see the term "interpolated" associated with DPI, that means the scanner software is creating extra pixels in between the ones that are scanned. The quality of the resolution has not only to do with DPI, but the quality of the optics used to make the lens and the brightness of the light source.

Other Considerations

Speed: Most scanners will show scan speed, in seconds per page, for both black and white and color. However, the speed of a scan is determined by several variables, including the type of computer you have, the DPI setting, and the complexity of the material you are scanning. Nonetheless, you'll see specifications like "20 secs/page (color)," which is about average for a good scanner. And unless you plan on doing a lot of scanning, a few seconds here or there probably isn't a big deal.

Negatives: One useful feature you might consider when buying a scanner is whether or not it has the ability to scan negatives and/or slides. Since negatives contain more detail than prints, you can scan them at 3000 DPI and then enlarge them with great results. So if you're planning to make enlargements of your pictures, scanning the negatives will give you a higher quality than simply scanning the prints. However, flatbed scanners are not well-suited for negative or slide scanning and you should be wary of claims from some manufacturers to that effect. Negative scanners start at around US$500, and if you decide to invest in one, you will also need specialized software to manage the scans. Rather than spend the money on expensive equipment, you can always get negatives scanned at a photo lab, which will put them on a CD for you.

Software: While most scanners include some type of software to manage the scanning operations, some bundles are better than others. For instance, with the HP scanner I bought there is the HP scanning software as well as a Photoshop lite program. So everything I need to scan and edit pictures was right in the box. Software bundles have gotten even more elaborate in the last few years, and now they often come with several different types of image editing programs.

Interface: The interface between your computer and scanner will probably be USB on either Mac or PC. You will sometimes see SCSI or Firewire connectors, but these are the exception.

Size: While most flatbed scanners are letter size, 8.5 by 11 inches, you can also find legal size scanners, at 11 by 14 inches.

Whatever you buy, be sure that it has all, or most of the features that you want. But what features *do* you want? Before you start a scanning project, there are two questions that you need to answer: What type of image am I scanning? How do I want so use it?

Scanning

Preparation

The way you scan an image depends very much on what you plan to do with the digital image. One of the first things you need to decide is whether the image is bound for print or the Web. Since the image pixels on a computer screen are bigger than the image pixels on a printer, you need to take this into account when scanning. That is, digital images that look good on a screen may look much worse when printed out on a printer.

So, if you are planning on scanning a picture from a magazine and posting it on the Web, or sending it over email, you can get away with a simple 72 DPI scan (which for a long time was the standard is the resolution of most monitors, but they're improving all the time — if you need to have exact control over the size of an image, you should scan with the exact resolution of your audience's monitor in mind). But if you want to scan that same image and print it out on a color printer, then it's worth scanning at a higher resolution, maybe something around 200 to 300 dpi. The tradeoff, of course, is file size. A 72 DPI scan might be only 100 KB, whereas a 6-inch by 4-inch photo scanned at 300 DPI can produce a 15 to 20 MB image. And if your scanner handles slides or negatives and you want to zoom in on a specific portion of an image, you may want to go as high as 2400 DPI! Then you're talking about Godzilla-sized files, which are measured in GB bytes.

Memory cost of images

Large images consume large memory and make our computers struggle. Memory cost for an image is computed from the image size.

For a 6x4 inch image at 150 dpi, the image size is calculated as:

(6 inches × 150 dpi) × (4 inches × 150 dpi) = 900 × 600 pixels

900 × 600 pixels is 900 × 600 = 540,000 pixels.

The memory cost for this RGB color image is:

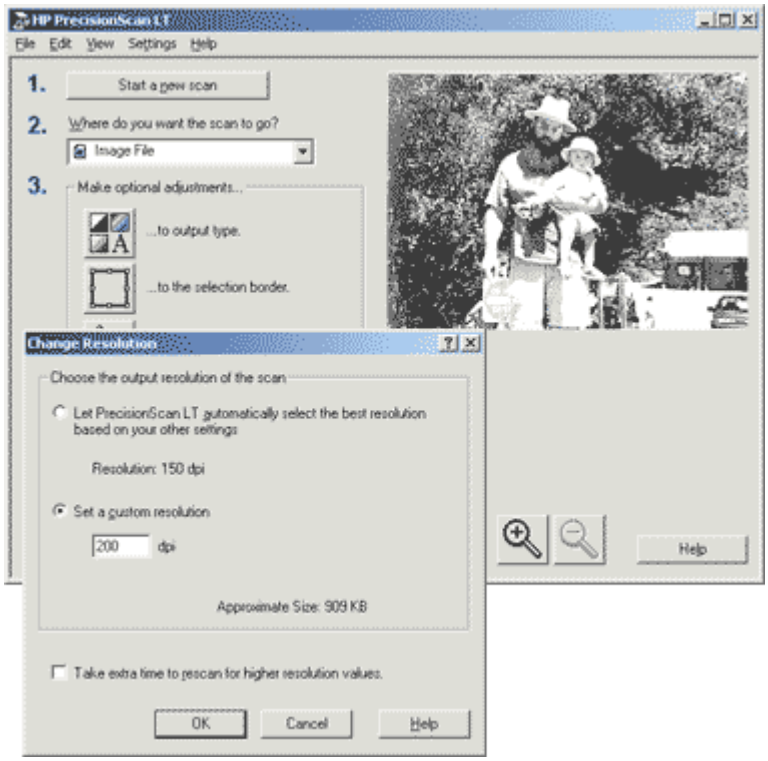
900 × 600 × 3 = 1.6 million bytes.

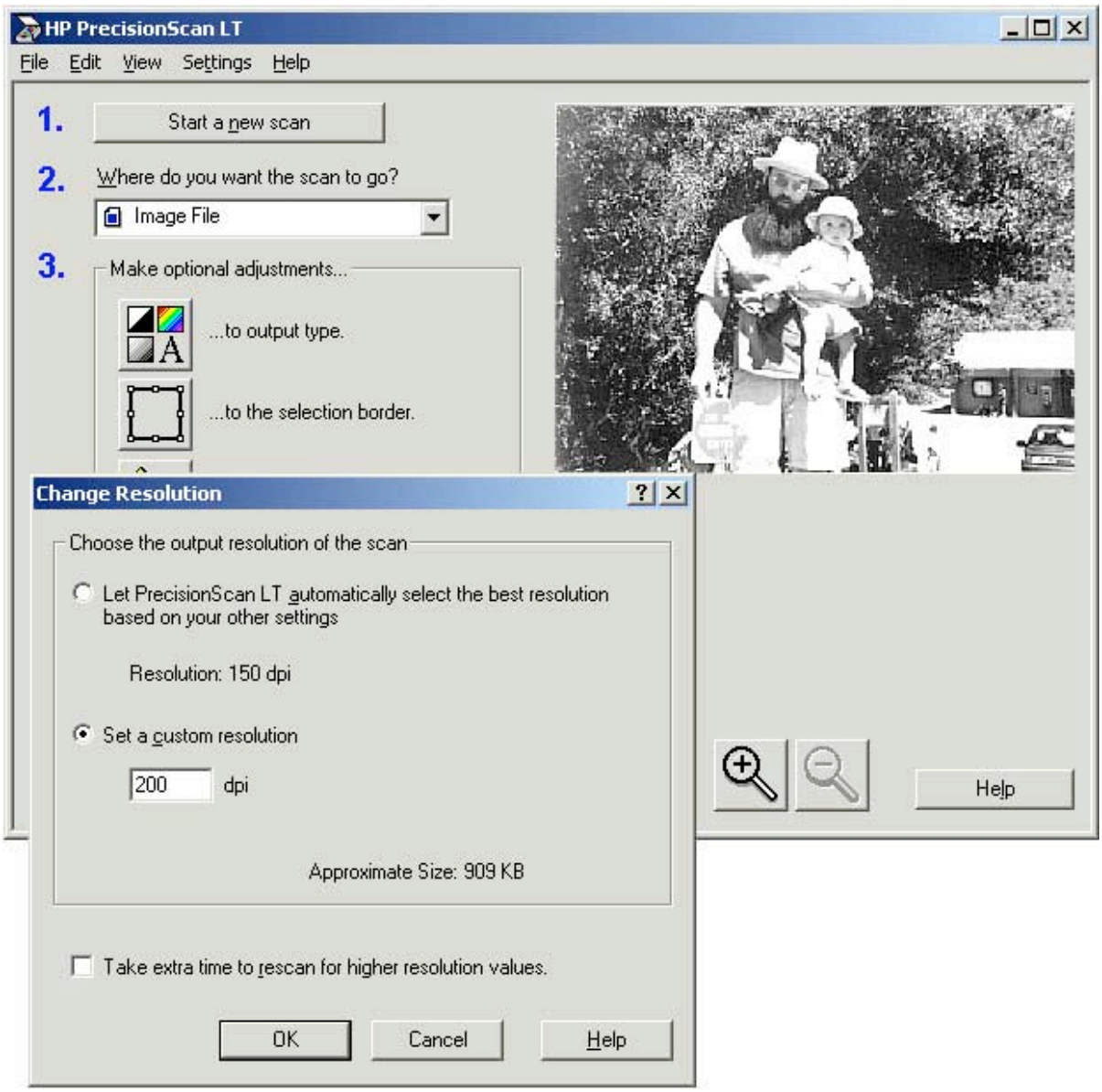
The last "× 3" is for 3 bytes of RGB color information per pixel for 24 bit color (3 RGB values per pixel, one 8-bit byte for each RGB value, which totals 24 bit color).

Different color modes have different size values, as shown below:

Image Type	Bytes per pixel
1 bit Line art	$^1/_8$ byte per pixel (1 bit per pixel, 8 bits per byte)
8 bit Grayscale	1 byte per pixel
16 bit Grayscale	2 bytes per pixel
24 bit RGB	3 bytes per pixel Most common for photos, for example JPG
32 bit CMYK	4 bytes per pixel For Prepress
48 bit RGB	6 bytes per pixel

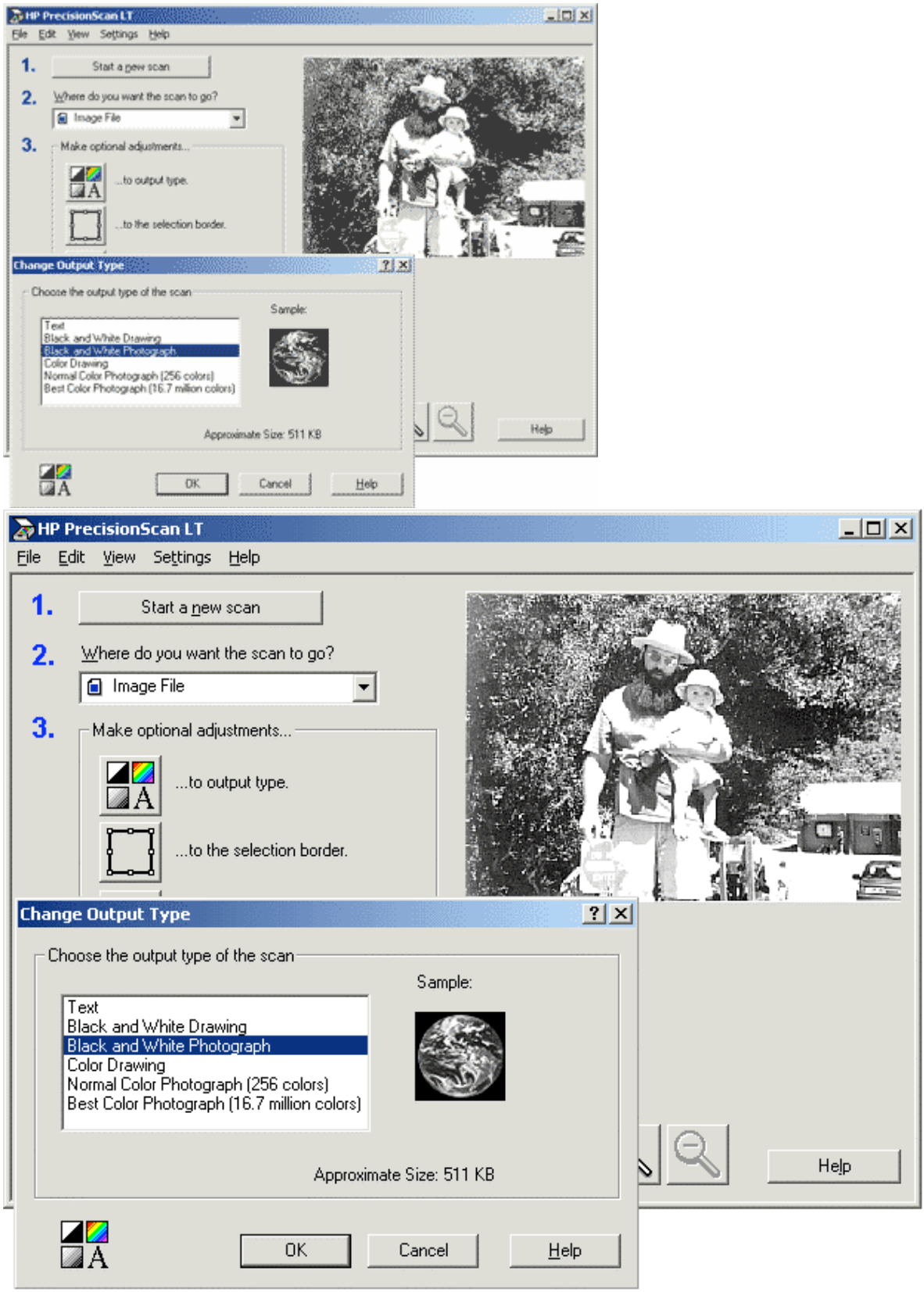
When you first scan an image, your best bet is to start with a low resolution setting and work your way up. The default resolution might be 150 DPI, and that's not a bad place to start. If you want to use your scans for both printing and the Web, you will want to scan them at two different resolutions, at least. For computers, images scanned at 72 to 100 DPI will look fine. For printing on a color inkjet, start at 200 DPI and work your way up. Your scanning software should give you a box where you can enter a specific DPI setting.


HP PrecisionScan LT
File Edit View Settings Help
1.
Start a new scan
2.
Where do you want the scan to go?
Image File
3.
Make optional adjustments...
...to output type.
...to the selection border.
Help
Change Resolution
Choose the output resolution of the scan
Let PrecisionScan LT automatically select the best resolution based on your other settings
Resolution: 150 dpi
Set a custom resolution
200
dpi
Approximate Size: 909 KB
Take extra time to rescan for higher resolution values.
OK
Cancel
Help



Go ahead and select the resolution. (Can't quite see what's going on here? Take a closer look.)

The next important setting you will make is the "output type," which describes what type of image you're scanning. These settings are pretty intuitive, and you will find that "black and white photograph" and "best color photograph" will be the best choice for most of your prints.



(Here's a closer look.)

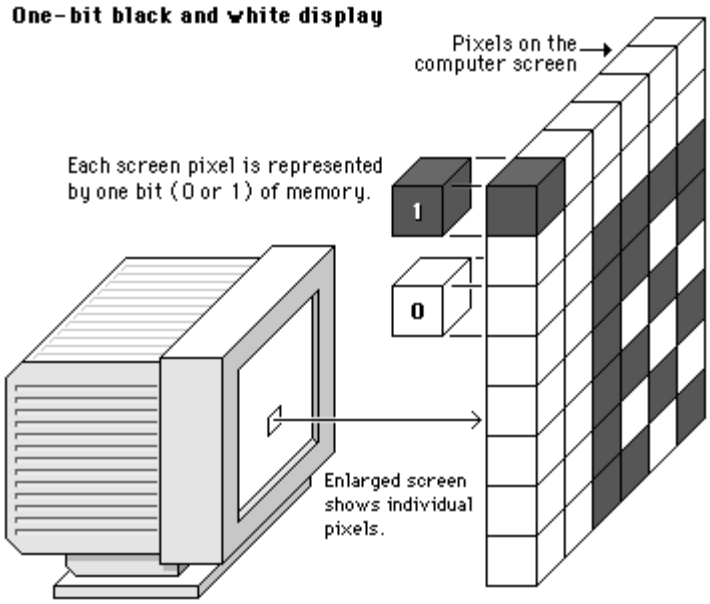
Once you've lined up all the pictures you want to scan, and decided what resolution to use, it's a good idea to close out of other applications on your computer. The more RAM the computer can use, the faster the scans. Likewise, you should make sure there is enough room on your hard drive to store the scanned images; you'll be surprised at how fast these files fill up your drive. It's also smart to get in the habit of trashing scans that you don't plan to save.

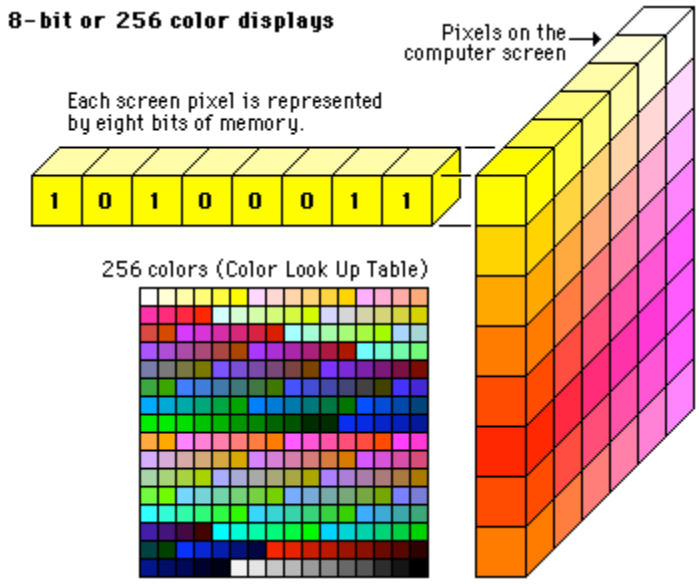
Another way to save space and time is to crop pictures before you scan them. Most any scanning software will give you a tool to select a specific area to scan, usually by dragging your mouse around the portion of the picture you want.

After setting your resolution and output type, make sure that your print is lined up correctly on the scanner bed. A preview screen in the scanner application will show you how the image looks. Once the image is situated correctly, simply press the scan button and wait for the sensors to do their work. Depending on the resolution setting, a scan will usually take anywhere from 10 to 30 seconds.

As always, please consult the manual if you're having trouble with the scanner or software.

How Scanning Works. Scanning is a process of sampling and recording the dots and color values that make up an image. The image is scanned and sampled line by line (rasterization) and the results are described as a bitmap file. Images are typically scanned at resolutions between 50 to 1200 Dots Per Inch (DPI). Bitmap files are commonly used in paint programs or image processing programs like *PhotoShop*.





Saving and Editing Images

Scanning is really a two-part process. The first part involves setting the resolution and output, and making the scan. The second part involves saving and editing the file

Most scanner software will let you adjust the color, contrast, and other aspects of your image while scanning. If you have a copy of Photoshop, or another good image editing program, don't bother adjusting these elements with the scanner software. Graphics editing apps give you more control over the process and generally produce better, clearer results. If you really want to streamline your scanning, you can configure Photoshop to act as your scanning software. However, it's just as easy to use the scanner software and save the image to Photoshop as a raw graphics file.

After you scan an image, you need to decide how to save it. When you pull up the save menu, you'll see a choice of about ten file formats, everything from BMP to JPEG to TIFF. These formats break down into two basic groups: compressed and uncompressed ("lossless" in graphics jargon).

When you save a file in a uncompressed format — TIFF, BMP, PCX — it won't discard any of the data during the encoding. The advantage here is that you retain as much information as possible about the scanned image, and can compress it later, after you've adjusted its contrast, brightness, etc. The disadvantage of these formats is that they create very large files.

If you like the way a scan looks right off the bat, you may want to just go ahead and compress it for the Web. In that case, JPEG and GIF are the two most widely used formats. The upside is that JPEG and GIF create very small files. If an image is 4 MB to begin with, JPEG will compress it to around 50 KB with only minimal loss in quality.

Remember, though, once you save a file in a compressed format, it will discard some of the image data, never to be seen again. The best course of action is to first save

your file in a uncompressed format, and then after editing, save it in a compressed format (and consider doing a "Save As" and save the original scan until your sure you don't need it anymore). In other words, do not compress the image until you are finished working on it, because every time it is compressed the quality is diminished.

I usually save files in BMP or TIFF and then import them into Photoshop. From there, you can manipulate the image using the sharpen filters, contrast, image sizing, touch-up, color balance, etc. When you have the image all spiffed up, go to the save menu and choose the format you want to save it with.

For images bound for a website, try to keep them under 40 KB. Web pages that have images bigger than this take a long time to load. If you're going to send a JPEG over email, something on the order of 60 to 100 KB will work.

If you compress a file with JPEG or GIF and find that it's still too big, go back to the original file, reduce its size (dimensions), and then resave it. If you do this a few times, you'll begin to see the relationship between image size and file size. Photoshop also features a "Save for Web" feature, which let you preview your image under a variety of compressions.

In general, JPEGs give you larger images and smaller files with better colors. So photographs are best saved as JPEG; logos or images with lots of flat color can be saved as GIFs.

For images that you want to print out, you should save them as TIFF uncompressed or BMP and use a high DPI setting (300 and up) when scanning them. This way, you'll retain the maximum amount of information from the image, and file size isn't a concern.

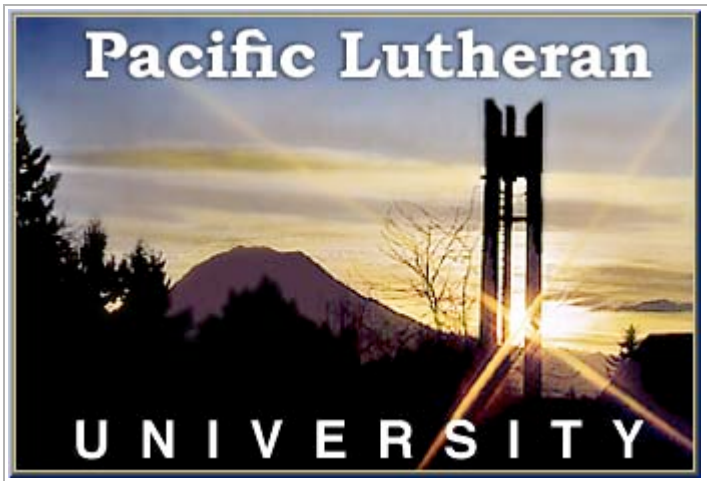
Now let's look at a quick case study from my own scanning adventures.

What Can Be Scanned?

Two dimensional opaque objects up to the size of a flatbed scanner (8 1/2 x 14) can be scanned in one pass, ususally up to legal size. Materials to be scanned include:

- photos
- artwork
- drawings
- text

Transparent materials, such as slides or negatives can be scanned with a special slide scanner or a flatbed scanner that has a transparency adaptor. Below are some samples of source materials, process used, and output options.

ource	Process	Output & example	
hotograph, egative, or ide	Scanning with flatbed ot slide scanner	`.jpg` file	
rtwork, ogo, graphic	Scanning	`.gif` file	
ext from age	Optical Character Recognition	`.txt` or `.doc` file	
ntire page	Adobe Capture	`.pdf` file	
anoramic hotographs	Scanning with slide or flatbed scanner; processing with QuickTime Panorama	`.mov` file	Click

File Formats

Digitized images can saved in a multitude of file formats. Many formats compress the image in some way to make the file size smaller. Several graphics formats provide reasonable compression and are cross-platform compatible. Not every format is compatible with every program.

Graphics compression can be lossless or lossy. Lossless compression looks for repetitive patterns, but doesn't change the image. Lossy compression also looks for repetitive patterns and averages parts of the image to make it smaller. All images should be saved with a lossless compression if further editing is required as averaged information can never be fully recovered.

One of the best cross platform graphics formats is TIFF (Tagged Image File

Format) which has both Mac & PC subformats but is usually readable on both and uses LZW compression (a lossless very efficient compression algorithm). For Internet use files should be saved in `.gif` or `.jpg` formats.

Output File Formats

Below is a list of some of the common file types for output.

File Extension	Description
`.txt`	ASCII text file; basic text without any formatting like bold or italic.
`.gif`	Graphic files often used for fancy text, graphics, logos, and animations. GIF (Graphics Interchange Format) is a 256 color or less representation of the original with no image alteration except for reduced number of colors. GIF compresses left to right, top to bottom and looks for repetitive patterns. GIF defaults to 256 colors, but some images will still look fine with as little as 3 bits of color information per pixel. Trying lower bit settings will make the file smaller but can severely distort color rendition. Convert images to Indexed color before saving as a GIF or use *Photoshop's* Gif89a export feature. You can always use the one step undo to get back to a true color image and try again if there is too much color distortion by trying a lower bit setting when using indexed color. Areas with subtle gradation usually don't index very well. Believe it or not selecting the graded area and adding about 10% noise to it before indexed color produces acceptable results.
`.jpg`	Image file often used for full-color photographs. JPEG files can be made from grayscale images or true color, 24 bit, 16.7 million color images. Files are compressed using a lossy compression algorithm. JPEG looks for areas of the image that are about the same and creates blocks that are reconstructed to create an image. JPEG has adjustable percentages of compression when saving and compresses very well. Photos compress with acceptable distortion via JPEG but text and lines usually suffer from jaggy & fuzzy edges. A new variation of JPEG is Progressive JPEG, which loads at successively higher resolutions (out of focus) until the entire file is loaded.
`.pdf`	Portable Document File for displaying facsimiles of documents including fonts, graphics, layouts, etc. Files are not digitally editable.

Following is a listing of file formats that can be imported for specific programs:

Web Browser

Web browser (such as *Netscape* and *Internet Explorer*) standard graphics formats are:

- **`.gif`** (normal, interlaced, transparent and animated)
    - for logos and buttons
- **`.jpg`** (normal and progressive)
    - for photographic continuous tone images

Microsoft *Word*

- **`.tif`** (PC and Mac)
- **`.gif`** (PC and Mac)
- **`.jpg`** standard, but not progressive
- **`.bmp`** (PC only)
- **`.pct`** or **`.pict`** (Mac only)

***Powerpoint* for Macintosh**

- **`.pct`** or **`.pict`**
- **`.tif`** (must be indexed color)

***Powerpoint* for the PC**

- **`.tif`**
- **`.gif`**
- **`.jpg`** standard, but not progressive

Adobe *Pagemaker*

- **`.tif`**
- **`.eps`**
- **`.jpg`**

Hyperstudio

- **`.pct`** or **`.pict`** indexed color @ 72dpi
    - usually image about 2 x 3 inch

What is OCR? Optical Character Recognition (OCR) is a process of scanning printed pages as images on a flatbed scanner and then using OCR software to recognize the letters as ASCII text. The OCR software has tools for both acquiring the image from a scanner and recognizing the text.

Ideal Source Material for OCR

OCR works best with originals or very clear copies and mono-spaced fonts like `Courier`. If you have choices, use the following source material:

- 12 point or greater font size.
- Black text on a white background.
- A clean copy; not a fuzzy multi-generation copy from a copy machine.
- Standard type font (Times, New Roman, etc.) Fancy fonts may not be recognized.
- Single column layout.

OCR Limitations

- Using text from a source with font size less than 12 points or from a fuzzy copy will result in more errors.
- Except for tab stops and paragraphs marks, MOST document formatting is lost during text scanning, (**Bold**, *Italic* & Underline are sometimes recognized).
- The output from a finished text scan will be a single column editable text file. This text file will always require spellchecking and proofreading as well as reformatting to desired final layout.
- Scanning plain text files or printouts from a spreadsheet usually works, but the text must be imported into a spreadsheet and reformatted to match the original.

What Source Material Doesn't Work Well for OCR?

- Forms (especially with boxes and check boxes)
- Very small text
- Multi-generation fuzzy or blurry copies from a copy machine
- Mathematical formulas
- Draft copies of documents with hand-written revisions
- Fancy text and unusual fonts
- Handwritten text

Adding Images to Your Web Pages

To insert an image into a web page the (**art_atwork.gif)** file, your basic image tag would look like this:

```
<img src="http://www.plu.edu/pics/art_atwork.gif>
```

To make the page load faster, include the width and height specifications of the image. This allows the browser to set aside space for the image when laying out the page so that text can be placed on the page while the image loads.

```
<img src="http://www.plu.edu/pics/art_atwork.gif" width="40"
                          height="40">
```

Inserting Images into *PowerPoint* Presentations

---

1. Open *PowerPoint* to the presentation you want to insert images into.
2. Navigate to the slide where you want to insert an image
3. From the **"Insert"** menu choose **"Picture"** or **"Picture/From File..."**
4. Find the file with the dialog box, select the file name, and click the **"Insert"** button.
5. When you click on the image to resize it, handles will appear on the corners of the image. Hold the "Shift" key down while you press your mouse button down over a handle. This will allow you to maintain the height/width proportions as you resize the image. Drag the handle diagonally to enlarge or reduce the image.

Image Editing

Using one of my photos as a guinea pig, let's go through a few of the basic Photoshop techniques you can use to sharpen up your pics (if you're using some other image-editing software, the principles should still apply).

The first shot I scanned was a black and white. I snapped a picture but it was slightly out of focus to begin with, and too bright as well. I thought I would scan it and then see how I could improve it in Photoshop.

Since I was scanning a black and white image, I needed to account for this in the scanner settings. With black and white images, the best setting to use is grayscale, also called "black and white photo." So after I set the scanner software to "black and white photo," I used the default DPI setting, which was 150.

After I scanned the image, I saved it as a TIFF file before sending it to Photoshop.

When I brought the image into Photoshop, I noticed that it was grainy and a bit washed out. My first instinct was to go back and rescan the picture at a higher DPI

setting. So I re-scanned the picture at 250 DPI and then imported that image into Photoshop as well. There wasn't much of a difference between the files, although the higher-DPI image showed slightly more contrast. The first TIFF turned out to be a 514 KB file, while the second was 1,373 KB. So the 100 DPI boost in the scan made the file almost three times as large!

Since I planned on turning this file into a JPEG and posting it on the Web, I wanted it to be as small as possible. For that reason, I decided to work with the 150 DPI image.

When I saved the image as a JPEG, I wasn't surprised to see that it looked about the same as the original picture, and was 180 KB to boot. I needed to make some changes. 

The original 180 KB version.

The first adjustment I made on the image was to fix the contrast. In the "Image:Adjust:Levels" menu, you can drag the arrow across until you get the contrast you like. Remember, in Photoshop you can "Ctrl-Z" anytime you don't like the effect of a change you make. The next adjustment I made was to fix the focus of the picture using the "unsharp mask" in the "Filter/Sharpen" menu. Both of these edits made the picture clearer and sharper, and I didn't want to push my luck much further.

Before I saved the picture, I decided to crop out the car that was on the right side of the frame. Not only was it a car, but it was some lame Mazda or something. It had to go, so I used the cropping tool to frame the image without it.

My final edit was to change the dimensions of the picture so I could reduce the final file size. In the "Image:Size" menu, I changed it to 600 pixels wide and 430 pixels tall.

In order to save the file as a JPEG, I went to the "File:Save" menu and chose JPEG at a "high quality" setting. When I went to look at the resulting file, I saw that it was still too big: 123 KB.

To reduce the file size even further, I went back to the original TIFF file (never trash your original until you're done editing and saving) and changed the dimensions to 400 pixels by 340 pixels. Then I resaved it as a JPEG again, and found that the resulting file was 25 KB.

This is the Web-sized version.

The final picture looked pretty good, and was certainly an improvement over the original print.