EYEBROW: THE FAQs ON SCANNERS

(From Practical Windows, April 1997)

Scanning Is on the Horizon

Straight answers to the 11 most frequently asked questions about scanner technology and today's top models.

by Lenny Bailes

What do you call a device that can shrink a stack of legal documents to an indexed collection the size of a 3.5-inch floppy disk? An electronic wizard? Try something more mundane: a scanner. And paperwork reduction isn't the only trick a little gizmo like this can show you.

Need a quick way to transform a printed letter into a form letter? Easy. Want to show off a great photo in your newsletter or Web page? No problem. Need color? It's going to cost you, but not nearly as much as it would have a few years ago. Once prohibitively expensive, scanners have entered the mainstream, turning up on desks all across the country. Whether you use Windows at home or in the office, a scanner can give your productivity a real boost. Here's what you need to know to buy the right model for your computing needs.

FAQ #1 How do scanners work?

A scanner is similar to a photocopier. A charge-coupled device (CCD) takes an electronic snapshot of a piece of paper, converting the light intensity of its content into digital information. You can store this information electronically as a disk file, dump it to a printer, or insert it as a bit-mapped image in a desktop-publishing program. You also can send scanned documents directly to a fax program or convert them with optical character-recognition (OCR) software into ASCII text that you can insert in your favorite word processor. The components used to accomplish this task are a photosensitive lens, a feed mechanism that lets you scan a specified area of the paper, and electronic logic that converts reflected light into a digital image.

Different design technologies allow scanners to record simple black and white, gray-scale or color representations of a reflected light source. The simplest scanners record an image in a black and white format; more sophisticated units can record various levels of gray shading or color.

Black-and-white scanners need only record intensity differences in a light source on a simple on/off basis. A dot on the page is recorded as either black or white. Gray-scale scanners look at the same image and convert reflected intensity into a monochromatic series of pixel shades. Like your video card, the scanner may support a varying number of gray levels--from four shades to over 16 million.

Older color scanners employed a three-pass mechanism to record hues, bathing a document successively in red, green, and blue light sources. Newer models use a more efficient one-pass technology. Color information can be registered through special filters on the CCD or with

specially designed trichromatic prisms.

The second important scanner component, the feed mechanism, delivers the document to be scanned to the photosensors. Running the photosensors over the paper is a mechanical process that can cause distortion of the resulting electronic image. Several different paper delivery designs are incorporated in three common scanner types. (See FAQ#2, below.)

The third important scanner component is the internal logic that translates the scanned information into a digital image. Depending upon the target application, the image may be scanned at different resolutions for faxing, preparation for OCR text conversion, or use in a desktop-publishing program. Algorithms built into the scanner smooth the overall texture of the image, weighing details and eliminating distortion caused by the mechanical scanning process.

FAQ#2 What are my options in scanners?

PC scanners ship in three popular configurations: hand-held, sheet-fed (or personal), and flatbed. Each design has its advantages and limitations. Hand-held scanners are the simplest and cheapest variety. They usually consist of a plastic mouse-like device that attaches to an extra serial or parallel port on your computer. By rolling the scanner over part of a sheet of paper, you capture its content as an electronic image. Hand-held scanners were very popular several years ago, due to the much higher prices attached to sheet-fed and personal models. Today, the price of more sophisticated scanners has come down and hand-held models are less attractive. Because a hand-held scanner depends upon manual manipulation to pass the photosensors over the paper, the images it captures are more prone to distortion. Hand-held scanners are also limited by their size to capturing a small portion of a page; usually, a 4- by 6-inch area is the maximum they can handle. Although you can sometimes perform successive scans to "stitch" multiple page segments together, the results may not be accurate enough for OCR software. The results of stitching bit-mapped images often prove even less satisfactory.

Older hand-held scanners were typically limited to a resolution of 200 dots per inch (dpi), and a 2-bit black-and-white color depth. This might be sufficient for capturing simple line-art or for transmitting a bit-mapped fax, but it's inadequate for OCR or for scanning more detailed art and photos. Some newer hand-held scanners can reproduce images in 16 or 256 grayscale shades at 400 dpi. Logitech's Scanman Color 2000 even records color images. But these models still suffer from the limitations of stitching and manual feeding.

Personal scanners are the next step up the ladder. With these devices, you insert a piece of paper in a slot, where mechanical rollers capture it and automatically draw it past the scanner's photosensors. Some companies offer "no footprint" sheet-fed scanners designed to ease the space crunch on your desktop. One such model, Visioneer's Paperport IX, incorporates a scanner into a PC keyboard; it can handle documents 8.5 inches wide by 30 inches long. The Storm Primax EasyPhoto Reader (see Photo 2, xxxx), on the other hand, is small enough to fit into a drive bay in your computer but can't accommodate a photo or document any larger than 4 inches wide by 6 inches long. (See the minireviews on pages xxx and xxx for further details on the Paperport IX and the Hewlett-Packard ScanJet 5p.) Personal scanners are faster than hand-held models and generally provide better results. They are a good choice for OCR conversion of a large number

of 8.5- by 11-inch pages, particularly when the manufacturer provides a multipage input tray like those available on most standard photocopiers. For a slightly higher price, the package also is likely to include software for managing documents and editing images. The limitation of a sheet-fed scanner is that it won't let you reproduce pages from bound books and magazines. For that, you need a flatbed model. Flatbed scanners are generally the most flexible and powerful models you can buy. But they occupy more desk space, and they cost more, although the price/technology gap is rapidly closing. Like a standard photocopier, a flatbed scanner lets you place individual pages, books, magazines, cloth patterns, and so on, face down on a transparent plexiglass surface. You close the cover and the photosensors automatically scan across the material to be copied. Because flatbed scanners are generally designed for high-end use, they are likely to support greater color depth and resolution (see FAQ#6). Flatbeds are generally perceived as more mechanically reliable than their hand-held and sheet-fed counterparts, simply because they require less user-input. Your hand can't zig where it should zag, and a crinkled or greasy piece of paper won't get caught between the rollers. In addition, the source document remains stationary while the photosensors slide up and down to scan it, so it won't slip if you insert glossy or flimsy paper.

FAQ #3 What's involved in setting up the various types of scanners?

Hand-held and sheet-fed scanners are the easiest to set up, because you generally can plug them right into a serial or parallel port on the back of your computer and start using them immediately. Some sheet-fed models let you attach your printer to the same port through a pass-through connector mounted on the scanner. This lets you continue printing without detaching your scanner. Flatbed scanners, designed for higher resolutions and deeper color-depths would be too slow if they used a serial/parallel port, so they usually are connected to a SCSI card or a card designed especially to work with a particular scanner. Usually, you can connect several SCSI devices in sequence (it's called <I>daisy-chaining<P>) and then connect to the SCSI card via a pass-through connector on the scanner. If you already have a SCSI card and SCSI devices such as a CD-ROM drive or hard disk installed on your computer, you probably can simply attach a scanner to the chain of devices. You'll need to make sure that you have the right cables and connectors if you plan to use an existing SCSI card. Your local computer electronics store should stock various SCSI converters and cables that bridge the different SCSI interfaces: 50-pin internal SCSI, 50-pin external SCSI-2, Centronix SCSI, and Wide SCSI-3. If you've previously set up the hardware and software for other SCSI devices, you'll simply need to plug in the scanner and install its software. If the scanner is your first SCSI accessory, you may have to do a little more work. You'll need to install the SCSI card inside your computer and get Win3.x or Win95 to recognize it. Generally, this isn't a difficult task. If the card is an older, non-PCI model, you'll need to make sure that your system has an available interrupt and port address, just as you would when installing a sound card. PCI SCSI cards generally configure themselves automatically. Under Win95, the plug-and-play capabilities of the operating system should automatically detect the presence of a new SCSI device and prompt you to insert the disk containing the scanner's software. Under Win3.x, you may need to manually install some scanner software--a so-called <I>driver<P>--in your [MONO]WINDOWS\SYSTEM<P> directory and load it in your [MONO]CONFIG.SYS<P> file. In the next year or so, though, the picture for Win95 setup may change considerably (see the sidebar "Coming Soon...," xxxxxxxx).

FAQ #4 What do I need to know about scanner software? @body text:Good scanner software may actually be the most important component of a scanner package. In the old days, scanning software made users enter a bunch of technical parameters to get decent results. Settings for resolution (the number of dots per inch), brightness/contrast, gamma adjustments, and so on were hidden in a maze of nested menus. Today, a good scanning application should do most of the work for you. The interface may automatically detect whether you've inserted a color, black-and-white, or gray-scale document. You should have to provide only minimal guidance--are you scanning a desktop-publishing image, creating a fax, or a doing an OCR conversion? The software should drive the scanner and transport the resultant image to an appropriate application on your Windows desktop. Some scanner packages now also include image-editing software, such as Adobe's Photoshop Limited Edition, Corel PhotoPaint, or Micrografx Picture Publisher. A good package also will include limited edition versions of professional OCR software, such as Caere Omnipage or Xerox Textbridge. If you intend to use your scanner for archiving paper documents to disk, look for a document manager such as Visioneer's Paperport or Microtek's PageWiz. Some document managers may include bundled OCR software. Many scanners now include a useful feature called TWAIN support. TWAIN is an interface standard that allows you to scan directly from and into common PC applications such as PageMaker, Photoshop, PhotoPaint, and Microsoft Imaging for Windows 95. If you want this added flexibility, you should determine whether the scanner you're considering ships with a TWAIN driver. Before buying the scanner, you should also audition its bundled software to see if the interface is intuitive and contains the features you need. It's also a good idea to ask about warranties, toll-free technical support, and the availability of driver updates.

FAQ #5 What format do scanners store images in, and how much disk space will those images consume?

Good scanning software should let you export images to all the common PC graphic formats: [MONO]TIF<P>, [MONO]BMP<P>, [MONO]PCX<P>, [MONO]GIF<P>, and [MONO]JPG<P>. A document manager may also let you save image libraries in a its own special format. All of these formats, except for [MONO]GIF<P> support images in as many as 16 million colors; [MONO]GIF<P> images are limited to only 256 color. How much space stored images occupy on your hard disk depends on the resolution, size, and color-depth of the scan, and the file format you choose. Some image formats, including [MONO]TIF<P>, [MONO]GIF<P>, and [MONO]JPG<P>, offer compression options that can greatly reduce the size of bit-mapped files. The table "Scans on File," provides representative file sizes for a 3- by 5-inch color photograph scanned at a number of different resolutions and color depths.

FAQ #6 If I want to scan color photos, what features should I look for?

To successfully scan color photographs, you should consider the quality of the CCD sensors built into the scanning device, as well as the resolution and color depth the scanner can support. The only way to really measure the quality a scanner's CCD is to scan the same photograph with different models and compare the results with the same output device (the same monitor/video card combination or printer for each scanner). The resolutions that a scanner supports may provide additional clues about its suitability for a given task. The resolution, the number of dots per inch in a scanned image, reflects the scanner's ability to detect and accurately represent tiny

details. Most common laser printers produce copy at a resolution of 300 dpi, and if you want to produce a photograph at the same resolution, your scanner should also perform 300 dpi scans. If you're scanning a photograph strictly for use in an on-screen presentation or web page, then a lower resolution of 75 dpi should be sufficient. Photographic scans intended for commercial publication should probably be done at a resolution of 600 to 1200 dpi. Most scanners that support color imaging are capable of recording images at a horizontal rate of 300 DPI. Some advertise the ability to scan 600 dpi vertically, but the vertical resolution is not as important, unless you intend to scale or manipulate the image later on. When you resize or alter an image, a higher vertical resolution may avoid distortion. You need to be careful when checking a scanner's rated resolution in dpi. The specifications for some models show two dpi capacities. The second, higher, dpi rating may be achieved through mathematical interpolation that doesn't reflect the scanner's true physical ability to detect light and create an electronic image. Interpolated scans at higher dpi ratings may turn out better than raw scans done at a lower dpi setting. However, interpolated scans are usually not as good as a raw scans done at a high resolution with a better lens. The second factor to keep in mind for color scans is the color-depth supported by the scanner. Color depth is measured in the number of shades that can be assigned to each pixel. One-bit color represents simple black and white, 2 [SUPERSCRIPT1] (two total colors). 8-bit color represents 2[superscript8], or 256 colors. An electronic image that reflects the same natural color palette the eye recognizes in a photograph requires a color depth of 24 bits, or 16 million possible colors. (24-bit graphics sometimes are referred to as Tru-Color graphics.) You may see ads that boast 30-bit internal color and 24-bit external color. This means that the raw scan of the image before it goes to the screen can contain several extra color levels for error-correction. The extra levels don't appear on the screen or printout, but they do present a larger palette from which the image-rendering software will choose the correct color scheme. A 30-bit scanner assists in tonal transformations, preserving very dark image details in shadowed areas of a photograph. If you plan to scan color photographs, you'll probably want to have a powerful image-editing program such as Adobe Photoshop or Corel Photopaint to adjust hues, saturation, brightness, and contrast. The scanner also should include some support for color-matching--calibrating the tones the scanner sees with the ones that your printer produces. Some personal scanners support color, but they generally aren't as sophisticated as flatbed models. (At this writing, Visioneer is planning a color version of their popular Paperport IX model.)

FAQ #7: Does the type of PC I have influence the type of scanner I should buy? @body text:The performance efficiency of serial, parallel and SCSI connections will vary with the CPU speed and the amount of memory on your computer. What works well on a 133MHz Pentium with 32MB of RAM may be too slow on a 486 DX2/66 with 8MB of RAM. One of the surest ways to measure your satisfaction with a scanner interface is to choose a typical document or photograph and measure the time it takes to run the scanner and pop the image up on your monitor.

FAQ #8: If I want to use my scanner primarily for OCR, what features should I look for?

To scan a document and convert it to editable text, you don't really need a color scanner. But you do need to consider resolution. The higher the resolution, the easier it will be for the OCR software to accurately convert a scan into text. Although some people consider 200 dpi (fine-fax quality) sufficient, you'll probably find that your OCR software makes fewer translation errors

with resolutions of 300 dpi or higher. Support for 256 gray levels may also be useful when the paper contains creases or has an uneven texture. OCR software works by scanning clustered groups of pixels on the page and making associations through lookup tables with alphanumeric characters and dictionary words. A full-featured OCR package will include automatic spell-checking of converted text or interactive features that let you correct the lookup associations. Such software also will prompt you when a pixel-group is unrecognized, allowing you to manually assign a letter or combination of letters to the pattern. A good OCR package should be able to convert bit-mapped images to text with a minimum of 98-percent accuracy. That may sound pretty good, but think about it. If 95 percent of 1000 characters on a page are converted accurately, that leaves 50 spelling mistakes to be corrected! Good OCR software should be also be able to detect differences in font size, bold and italic text, preserving these changes in standard word processor formats (Microsoft Word, WordPerfect, and Rich Text Format).

FAQ #9 Which footprint type is the best option for OCR? @body text:Most OCR programs can do ASCII translation directly from standard bit-mapped image formats, but the results of OCR conversion with a hand-held scanner are likely to be less satisfactory. When you stitch different portions of a page together with a hand-held scanner, you're likely to introduce variable spacing between words. That may confuse the OCR package. If you need to scan pages from a bound book, of course, you'll need a flatbed scanner rather than a personal scanner, unless you also have access to a flatbed photocopier. If you plan to scan a great many loose pages, you'll probably want to look for a scanner model that supports a multipage input tray (personal scanners) or an automated document feeder (flatbed models).

FAQ #10 What do I need to be able to file, duplicate, or fax the documents I scan?

You need a document manager to provide a filing system for your scanned images. Typically, such software offers a hierarchical tree of folders into which you can drag and drop files. A good document manager also will let you identify scanned images through thumbnail representations and keyword summaries attached to each file. Currently, Visioneer Paperport is the state of the art in document management software (see Photo 3). It can accept images directly from a number of scanners; its intuitive interface lets you drag thumbnails and drop them on icons that will fax a scanned document, perform OCR conversion, send an image to a printer or image-editing application, allow you to add annotations, or file the image away in a folder. Paperport is bundled with Visioneer's own personal scanners and a version of the program is included with most Hewlett-Packard scanners. At press time, Visioneer had just released a stand-alone version of this product called Paperport Deluxe. Paperport Deluxe can be used with any TWAIN-enabled scanner and includes more features. Paperport Deluxe automatically performs OCR translations of scanned documents in the background and indexes them so that you can do keyword searches. Its filing system can support as many as 10,000 nested folders, and you can automatically back up scanned images into compressed archive formats. Paperport Deluxe has special support for scanning in business cards and also includes Visioneer FormTyper. FormTyper converts scanned forms into OCR documents, so that you can fill in the blanks, sign the completed form, and re-fax it. Other document management packages bundled with popular scanners include Documagix Papermaster and Microtek's PageWizard and ScanWizard.

FAQ #11 What can I expect to pay?

Black-and-white hand-held scanners go for $99 to $200, hand-held gray-scale and color scanners come it at $200 to $300. Sheet-fed personal scanners generally fall in the $200 to $400 range, and flatbed models are $400 and up.[ENDBOX]

SIDEBAR 1 Visioneer Paperport IX @sidebar text:The perfect scanner for simple line-art, gray-scale photos, simple document faxing, and OCR conversion, the Paperport IX is a 101-key professional keyboard with the sheet-fed scanner built right in. It uses an external power supply, a standard keyboard connector, and a 9-pin serial port. (Visioneer has adapters for 25-pin serial port and parallel port connections.) The unit supports scans at 200 dpi (400 dpi, interpolated) and can produce as many as 256 gray levels. You just pop a photo, business card, or sheet of paper into feeder and within seconds the scanned image pops up in the Paperport document manager. From there, you can file the image in a folder, view it, save it to disk in any of the common bit-mapped and fax formats, convert it to text, fax it, or print it. Although the Paperport IX isn't a TWAIN-enabled scanner, the Paperport software supplied with the unit makes it easy to cut and paste documents into other applications. Or you can upgrade to Visioneer's new Paperport Deluxe ($99), which includes direct scanning support for such popular PC applications as Pagemaker and Photoshop. The unit that we tested can be hooked up to either a serial or a parallel port, but it doesn't support a pass-through port connection. Paperport IX, $299. Visioneer, 34800 Campus Drive, Fremont, CA 94555; 510-608-0300, fax: 510-608-0305; <B>www.visioneer.com<P>. System requirements: Win3.x or Win95, a 386 or later computer, 4MB RAM (8MB for OCR), 12MB of disk space, CD-ROM drive (3.5-inch disks available on request). --L.B.

CAPTION: The Visioneer Paperport IX is a 101-key professional keyboard with the sheet-fed scanner built right in.

SIDEBAR: Hewlett-Packard ScanJet 5P @sidebar text:An attractively designed, lightweight scanner, the ScanJet 5p works well with its own or a third-party SCSI connection. This inexpensive flatbed scanner supports 24-bit gray-scale or color scans at 300 dpi (1200 dpi, interpolated); it can handle pages as large as 8.5 by 11.7 inches. The scanner includes a 16-bit ISA SCSI card, cables, and the necessary software to configure it. The HP PictureScan software supplied with the unit can save images directly to disk, make copies on a printer, or dump the image into the bundled version of Visioneer Paperport. The image-editing software included with the package is Corel PhotoPaint 5.0. <I>Hewlett-Packard ScanJet 5P, $399. Hewlett-Packard Co., Direct Marketing Organization, P.O. Box 58959, Santa Clara, CA 95051; 800-722-6538; <B>www.hp.com<P>. System requirements: Win3.x or Win95, 486 PC or later, 8MB RAM, 20MB available disk space, and a CD-ROM drive (3.5-inch disks available on request). --L.B.

CAPTION: Well-designed yet inexpensive, the Hewlett-Packard ScanJet 5p flatbed produces 24-bit gray-scale or color scans.

SIDEBAR 3: COMING SOON... @sidebar text:A new kind of peripheral bus is coming to Win95. Microsoft and Intel have collaborated on a new standard called the Universal Serial Bus

(USB). The advantage of this interface, which should start appearing on computers in late 1997, is that you'll no longer have to open your computer's case to install new hardware devices. A single plug-in connector on the back of the computer will allow you to daisy-chain as many as 127 different peripherals, including keyboards, modems, printers, monitors and--of course--scanners. USB devices require no external power supplies, and all cooperate with a single port and interrupt assignment. No more hunting for unused connectors on the back of the machine and shuffling cards around to juggle interrupt/port resources. No USB scanners are available right now, and even if they were you'd also need a USB-compatible motherboard and special upgrade of the Win95 operating system to take advantage of them. But by next year, all this new technology should be in place. Another new wrinkle on the PC hardware scene is Intel's Pentium chip, with its multimedia extensions (MMX). MMX Pentium chips contain a new instruction set that lets specially programmed applications work more efficiently. Currently, Adobe's PhotoShop image-editing program is MMX-enabled. Running it with an MMX CPU will reduce the time it takes to resize, rotate, and manipulate large graphic images. --L.B.