IMAGE CAPTURE

Images must be captured from source documents or originals. Thereafter, they can be edited, stored, organized (cataloged or indexed), transmitted, and accessed. Images can be captured from books, magazines, photographs, paintings, slides, motion pictures, microfilm, microfiche, and many other types of source documents.

There are two main types of capture devices: scanners and digital cameras. Most scanners are similar to photocopiers and fax machines; the source documents come in contact with the capture device. A few expensive models—designated planetary or faceup scanners—are similar to planetary microfilming machines; the source documents do not come into contact with the capture device. Digital cameras, when mounted on a camera stand, are also like planetary microfilming machines; however, they have far less capacity than a digital scanner. Until recently, their advantage was the 32- to 42-bit color they offered, but that advantage has disappeared now that many flatbed scanners also offer that.

Scanners

Documents can be placed in a feeder tray and fed into the system using a workflow scanner (common when capturing images of correspondence, insurance policies, checks, credit card charge slips, and so on). They also can be scanned with a hand-held scanner moved over the source document, placed face down on a flatbed scanner, or placed faceup several feet below a scanner.

The first type is not suitable for libraries and archives because the source documents are often adversely affected by the data capture. For example, books and magazines have to be unbound and light-sensitive photographs, manuscripts, and maps may be damaged because the source document is passed over a round drum. The second type is also unsuitable because the pressure of the hand-held scanner cannot be carefully controlled.

Flatbed Scanners

A flatbed scanner—the type most often used by libraries and archives to capture images of bound books, manuscripts, photographs, and so on—is literally a flat glass bed, quite similar to that of a copy machine on which the source document is placed face down and covered. Flatbed scanners require the user to lift, rotate, and place the documents face down. Although not time-consuming when unbound sheets or manuscripts are to be scanned, it is time-consuming when bound books or journals are to be scanned; and there is also the risk that fragile bound materials will be damaged by pressing down on the spine to eliminate a shadow.

Flatbed scanners range in price from inexpensive desktop models costing under $100 and scanning a document in 30 or more seconds to fast machines costing up to $2,500 and scanning a document in as few as six seconds. The lowest price scanners are consumer models; the highest price ones are designed to meet the needs of companies and organizations. The latter often are designated "professional" models to distinguish them from the consumer products.

Until recently, flatbed scanners tended to be limited to 200, 400, or 600 dpi. 200 dpi is comparable with the resolution achieved with fax machines; 400 dpi is comparable with high-quality photocopiers; and 600 dpi is comparable with the quality of a high-resolution laser printer. Beginning in 1998 flatbed scanners priced higher than $400 offered a minimum of 600 dpi and some offered as high as 3,200 dpi. In early 2000 even $100 flatbed scanners were offering 600 dpi.

Most flatbed scanners allow setting of the resolution at time of capture so the operator can decide the appropriate resolution for a particular source document. This decision reflects the desired balance between scanning time and quality.

A scanner should be TWAIN-compliant. TWAIN is the standard communication protocol between a capture device and imaging software. An older proprietary communication protocol known as ISIS is still offered as an option on some scanners.

This report focuses on the professional models priced at $400 to $2,500 as these are the ones most often used in libraries and archives.

Planetary or Face-up Scanners

Planetary or faceup scanners are more expensive than flatbed scanners, typically $15,000 and up. In the past, they have been purchased by libraries and archives because they were superior to flatbed scanners in resolution, control over the capture process, and protection of the source documents.

Now that many flatbed scanners offer resolution of 1,200 dpi and more, resolution has ceased to be a reason for choosing a planetary scanner.

Planetary scanners continue to be superior to flatbed scanners because they provide greater control. For example, it is possible to adjust the camera so that source documents of different sizes are captured at a uniform size. With flatbed scanners, this type of adjustment is usually made later in the editing process. However, flatbed scanners with features similar to those of planetary scanners are becoming more common.

The most compelling reason to select a planetary scanner instead of a flatbed scanner is that the former provides greater protection for the source documents. The planetary scanner not only does not touch the source document, but a bound volume can be gently opened without pressing down on the spine. The digital scanner is mounted on a stand above the source document. The distance to the source document can be controlled manually or with a motor. Lights mounted to the sides of the scanner eliminate any shadows on the source document, including the gutter shadows of bound items. Fragile manuscripts and bound materials are best captured on this type of scanner.

Digital Cameras

Digital cameras—used primarily for digitizing color originals—are similar to traditional cameras, except that rather than exposing film, they expose a Charge-Coupled Device (CCD) array to light and produce digital output. A digital camera does not come into contact with the source document. Although digital cameras are meant to be hand-held, they are rarely hand-held when used in image capture. Instead, the camera is mounted on a camera stand. A light source is mounted above and to the side of the camera, and below the platen. A digital camera does not provide as much control as a digital scanner; most adjustments must be made during subsequent editing of the images. Digital cameras typically are priced at under $900, but the stand adds several hundred dollars more.

Digital cameras are used primarily by professional photographers and graphic artists, rather than by libraries and archives. However, a library that does not have a planetary scanner may wish to consider purchasing a digital camera and camera stand for capturing images from a particularly fragile source document.

Ideal Capture Device

The ideal capture device for libraries and archives is a planetary or faceup digital scanner because:

- It allows scanning of bound materials in the faceup position using an adjustable cradle to provide nonstressful support.
- Most correct page curvature and line skewing, and remove shadowing so even information near the binding can be read clearly.
- Most also have automatic edge detection so borders and the area beyond the page are masked out.
- Almost all set the focus and exposure automatically, although these also can be set manually.
- With few exceptions, two facing pages can be scanned at once.

Unfortunately, planetary or faceup scanners are more costly than flatbed scanners. A library or archive that cannot afford such a scanner has three options: outsource the work when fragile source documents are involved; exercise extraordinary care in capturing the source documents with a flatbed scanner; or purchase a digital camera and camera stand for low-volume use.

Ideal Capture Resolution

As recently as 1997, libraries were capturing images at 200 and 300 dpi. In many cases, it has been necessary to re-scan the images now that access devices, both monitors and printers, support resolution greater than 300 dpi. Given the high cost of labor and the relatively low cost of the initial storage media on which the images are stored, it makes sense to capture a high-quality image initially.

The ideal capture choice is 24-bit color and 1,200 dpi for most libraries and archives because that capture choice offers a high-quality image at moderate cost. The image looks more photo-realistic even if the original is not in color. Once captured in 24-bit color, an image can be copied to grayscale for use, with the 24-bit color retained as an archival copy. If the archival images are at 1,200 dpi, it should not be necessary to recapture the images at a later time to improve the resolution of a service image to the level of current access equipment, including monitors and printers. Typically, low-cost image capture devices offer higher resolution than low-cost monitors and printers, but the resolution of the latter is constantly getting better.

The minimum a library or archives should consider is 24-bit color and 600 dpi because current access equipment supports that level at modest cost.

Image Formats

As mentioned previously, four common image formats transfer to almost any platform or software system: TIFF, JPEG, GIF, and PDF. The decision as to which is to be used should be made before selecting a capture device because not all support all file formats.

TIFF

TIFF (Tagged Image File Format) files are widely used for the archival or master copy of a file. Once saved, a TIFF file can be retrieved and read by a computer with a

completely different hardware and software system. TIFF images are easy to manipulate. When compressed, there is no loss of information (lossless compression).

JPEG

JPEG (Joint Photographic Experts Group) files are widely used for Web viewing and transfer through systems that have space restrictions. The compression—which typically is somewhere between 10:1 and 40:1—is lossy, meaning that as a file is compressed, it loses bits of information. This does not mean, however, that the image markedly decreases in quality. If it has been scanned at 24-bit color, the loss is not visible to the human eye. JPEG images cannot be enlarged without loss of quality. Although JPEG images are not popular for archiving, they are popular as service copies. JPEG is used to distribute images on the Web. All Web browsers support it.

GIF

GIF (Graphic Interchange Format) files are an older format developed by CompuServe Inc. that is limited to 256 colors (8-bit color). Like TIFFs, GIFs use a lossless compression format, but they require less storage space. It should be avoided.

PDF

Although all the foregoing image formats are based on international standards, PDF (Portable Document Format) is a proprietary format developed by Adobe. PDF is widely used on the Web because of the attractive features of the Adobe Acrobat reader, so it is becoming a popular format for imaging systems.

Despite its popularity, it should be selected only after careful evaluation of its merits because it is not a true standard.

Several other proprietary image formats, among them Microsoft's BMP and Apple's PICT, should also be avoided.

The most common approach to image capture is to create both TIFF and JPEG files at the time of scanning, the former for archival use and the latter for service use. Alternately, TIFF files can be created at time of scanning and JPEG copies made from them as part of the subsequent editing process.

Popular Flatbed Scanners

Flatbed scanners are available in more than 150 models, but only a dozen or so have been successfully used in libraries and archives. The most widely used to date are the models of the Hewlett-Packard 3C "professional" series. Although no longer being manufactured, the company currently offers many models with similar features, especially those numbered 6300. Several other vendors have so-called professional models—a term that differentiates them from consumer models. All the models discussed are TWAIN-compliant.

HP ScanJet 6300C

At a price of $400, this model is an excellent investment for an organization committed to capturing images of text in a minimum of 24-bit color at 1,200 dpi with a tight budget. The 6300C is actually rated at 36 bits. It scans quickly, a 4-by-6-

inch color photograph in nine seconds and a typical page of text in one-third that time. Both a USB and SCSI interface area available to Windows 95/98, Windows NT 4.0, and Mac OS 8.5 and above. Basic editing tools are incorporated, specifically Adobe PhotoDeluxe Business Edition 1.1, a basic image-editing package. You must upgrade to Adobe PhotoShop V5.5 or 6 to enhance photographs or old manuscripts. Other models in the 6300 series include automatic document feeding (not suitable for libraries and archives) and special Web site creation software.

Canon CanoScan FB 1200S

This 36-bit color, 1,200-dpi model is priced at $500 but is often discounted to $300. It is an excellent choice if imaging is to include color photographs and old manuscripts. Like most models priced under $1,000, it has a maximum scan width of 8.5 inches and a maximum scan height of 11.7 inches, therefore, it may not be suitable for some manuscripts, and certainly not for maps. Scanning speed is comparable with the other models priced under $1,000: about nine seconds for a 4-by-6-inch color photograph and one-third that for a typical page of text. It is a SCSI-based device that can be used with all Windows-based PCs. It has an optional film adapter unit for transparencies and negatives ($280). The model is loaded with Adobe PhotoShop 5.5 editing software, a choice suitable for any image-editing task.

For more information, go to the manufacturers' Web sites:
www.hp.com
www.canon.com
www.microtek.com
www.epson.com
www.fujitsu.com

Microtek ScanMaker X12USL

Introduced in the third quarter of 2000, the ScanMaker X12USL is a professional model. Pricing had not yet stabilized at the time this report went to press, but it was more than $1,000. The scanner provides 2,400-by-1,200-dpi optical resolution, full-in-and-out 42-bit color, and USB connections. It includes on-the-fly image compression for large files, a LightLid 35 transparency adapter for 35mm slides and filmstrips, an 8.5-by-11.7-inch scan area, and a Zero Reflection Design technology to eliminate reflections from the source documents. The software applies advanced artificial intelligence techniques to automatically determine the size of a document and adjust the color, brightness, and contrast. It includes Adobe Photoshop LE 5.0 for image editing and Trellix Web publishing software.

Epson Expression 1600 Professional FireWire

This top-of-the-line scanner model is rated at 36-bit color, and it offers a resolution of 1,600 dpi in one dimension and 3,200 dpi in the other. Although it costs $1,400, you gain not only the greater resolution, but also a faster scanning speed: four seconds for a 4x6-inch picture or less than two seconds for a typical page of text. Both SCSI and USB interface connections are offered, as is IEEE 1394 (FireWire). The USB interface is substantially slower than the other two interfaces. The unit is configured with Adobe PhotoShop 5.5 editing software under Windows 98. FireWire has a Windows 2000 option.

Fujitsu ScanPartner 600C

The ScanPartner 600C, a 24-bit machine, offers 600-by-1200-line resolution of images up to 8.5 by 14 inches in size. It uses a SCSI-2 interface. It offers only PDF, so it is well-suited for getting documents online, but not for other applications. It is TWAIN compliant. There is an ISIS option. The unit has an automatic document feeder—a feature most libraries and archives will probably not use. The software includes Photoshop LE, PageMill (for Web creation), and Acrobat (for PDF file creation). The list price is $1,795.

Popular Faceup or Planetary Scanners

The faceup or planetary scanner most widely used in libraries and archives is the Minolta PS 3000 Publication Scanner. It can scan up to 11 by 17 inches at a resolution of 200, 300, or 400 dpi. The scan mode is monochrome with 256 gray scales (that is, 8-bit). The copy can be reduced by 50% or magnified up to 200%. The minimum time to scan a printed page is about six seconds. Images can be exported in TIFF or JPEG format via a SCSI interface. The data transmission speed is 1.7 Mb per second. The scanner is most appropriately used on printed text and should not be used for colored illustrations. The price is about $17,000.

The company has recently improved the PS 3000 and developed a complete turnkey system designated the MI3MS 3000 Plus.

Despite its popularity, the limited resolution makes this scanner a less desirable choice than some other digital scanners.

A somewhat more suitable scanner than the Minolta PS-3000 is the PULNIX MFCS-50. The major advantages over the Minolta model is its 24-by-32-inch scanning area and the ability to capture images at up to 600 dpi. It offers a choice of black and white or bi-tonal (1-bit), 4-bit (16 gray levels), and 8-bit (256 gray levels) scanning. If the material scanned is more than 10 by 17 inches, the maximum is 400 dpi. You can scan only the left page, right page, top half page, or bottom half page. There is no cradle but the special lighting avoids shadows. Brightness, contrast, and threshold settings are automatically set or can be adjusted. The camera head height, focus, and aperture are easy to adjust according to guidelines provided for optimal imaging. TIFF images are transferred via SCSI-2 to a host computer. The list price is $17,000. Although the MFCS-50 can be operated using a variety of scanning software, OMNISCAN scanning software has been designed specifically to work with it. It is priced at $4,000.

The widest range of faceup or planetary scanners is produced by Zeutschel GmbH of Germany and distributed in North America by Crowley Micrographics. The Zeutschel OMNISCAN 7000 AO is the top of the line. Both a table and cradle are available. It scans at up to 800 dpi. Electronic image correction of the book curve is optional, and essential for libraries. The scan head can cover an area of about 33.5 by 47.5 inches. Both automatic and manual modes are available. The output formats are bi-tonal black and white 9 (1-bit), photo mode (8-bit color), and 16/256 gray scales (4-bit and 8-bit). Scanning speed is as fast as 3.2 seconds. The image formats supported are TIFF, JPEG, GIF, and PDF. It has optional enhancement software, including contrast enhancement, rotation, despeckel, deskew, crop, mask, and so on. There is a SCSI-2 interface, and both ISIS and TWAIN are supported. The scanner can be used with black-and-white text, illustrations, maps, manuscripts, many photographs, and three-dimensional objects. Prices vary a lot from distributor to distributor, but usually are in excess of $30,000.

The OMNISCAN 6000 Color is similar to the 7000, but uses digital color filtering technology to achieve true color. The size of originals is limited to 24 by 33.5 inches and the maximum dpi is 600. Scan speed is a minimum of five seconds. The scanner is the most appropriate when the source documents contain colored illustrations.

Yet another option is the OMNISCAN 5000. It is similar to the 7000 but is limited to 600 dpi and a scanning area of 17.4 by 25.4 inches. The electronic book curve correction is standard on the alternate model OS 5100 TT. The interface is a SCSI-3 interface, and both TWAIN and ISIS are supported. It is the fastest in the product line at a minimum of two seconds. It is recommended primarily for printed text and manuscripts, not photographs. The list price of the 5100 is $29,500, but it often is discounted 20% or more, especially when ordered directly from Crowley Micrographics, the importer.

For more information, go to the manufacturers' Web sites:
www.minolta.com
www.pulnix.com
and for all products distributed by Crowley Micrographics use www.crowleymicrographics.com.

Finally, the Zeutschel Omnia 300/301 Hybrid offers both digital scanning and microfilming. A microprocessor-controlled hybrid camera system is designed to scan or microfilm onto 35/16mm roll film simultaneously all kinds of documents and books at up to 800 dpi. The scanning speed is 6.5 seconds at 800 dpi and 3.2 seconds at 400 dpi. There is a SCSI-3 scanner interface. You can chose between a TWAIN or ISIS scanner driver. All standard outputs are supported, including TIFF, JPEG, and PDF. Depending on the options selected, the price can go as high as $65,000.

Popular Microform Scanners

The leading manufacturer of scanners that capture images from microfilm and microfiche is Canon. The Canon MS-800 handles all types of microfilm and fiche, although the carriers have to be changed when going between fiche and film. Both 16mm and 35mm microfilm on open reels or in cartridges are accommodated, as are microfiche, jackets, and aperture cards. Both negative and positive film and fiche, and silver and diazo film and fiche are accommodated. Its main advantage is its 600-dpi resolution. It offers multiple scanning modes: fine, photo, grayscale. File formats include TIFF and JPEG. There is a SCSI-II interface. The base price is under $500, but its many optional attachments can significantly increase the price.

Another manufacturer of microfilm and microfiche scanners is Mekel, a firm also represented by Crowley Micrographics. The Mekel 500 Microfilm Scanner can scan more than 100 frames per minute, including 16mm or 35mm microfilm; positive or negative; silver halide, diazo, vesicular, thin or normal; open reels or cartridges; and 100-, 215-, or 1,000-foot reels. Maximum resolution is 400 dpi. Images can be created as bi-tonal or 256 grayscale. The outputs are CCITT 4 compression (that used in late-model fax machines) for bi-tonal and JPEG when grayscale is employed. The price is about $2,500.

The Mekel 565 Microfiche Scanner is available in bi-tonal and 256 gray scale models. At 24x the resolution is 100 to 400 dpi, at 48x it is 100 to 300 dpi. The fiche can be jackets, COM, A.B. Dick, or Microx type, positive or negative; silver, diazo, or vesicular; cine or comic mode. Up to 75 fiche or jackets can be stacked in the auto load tray. A prescan feature automatically determines image location and size. There is automatic density compensation, frame edge detection, deskew, crop, rotate, and auto contrast. The output is CCITT 4 compression or JPEG, the latter when grayscale is employed. The price is about $2,500.

Digital Cameras

Digital cameras are most widely used for scanning photographs, but are also suitable for text. Most systems are not offered as turnkey products, but as individual components. Although a digital camera and stand may cost as little as $1,500, it can cost many thousands. Typical of a high-end system is that installed at the National Archives of Canada. The configuration includes a high-quality Eikonix EC850 planetary photodigitizing camera, a PC with a VISTA video processor board, VGA controller, a 19-inch 1,024x768 image monitor, a 14-inch data monitor, and Laserdrive 810 5.25-inch digital WORM optical disk drive. Illumination is provided by two 1 Kilowatt Lowel Tota lights on stands, each equipped with a constant voltage regulator power supply. (The high-intensity quartz lamps that come with the Eikonix camera generate too much heat and provide uneven illumination and fluorescent lights flicker at 60 Hertz, causing the scanned image to have alternating bands of lighter and darker lines.)

To scan an image, the camera operator places an original on the illuminated bed of the photodigitizing camera and focuses the 55mm Nikon Macro lens using

electronic feedback. The thickness of a horizontal bar on the monitor varies as focus is adjusted, displaying the difference in intensity between adjacent pixels. Maximum sharpness is achieved when the focus bar appears thickest. Exposure, affecting image density and contrast, can be adjusted via the lens aperture or by increasing or decreasing the exposure time of each step in the scanning. In practice, the fastest scan rate is usually used and exposure adjusted by opening up or closing down the lens aperture.

The camera has a small linear array located at the film plane of the lens. As the array moves across the image, each pixel of the element emits an electrical signal proportional to the intensity of the light that strikes it. These electrical signals are digitized and stored in the memory of the video processor board. The video processor displays the stored data one scan line at a time that provides the camera operator with instantaneous feedback as an image is gradually built up. This feedback allows the operator to assess image quality, adjust the variables of focus, exposure, or lighting, and rescan the image. Once an entire image is scanned, the operator is prompted to accept or reject it. When a scanned image is accepted, it is stored in the video processor board's 4 Mb RAM buffer and the operator is prompted to enter image tagging information consisting of essential control data. This data includes the accession and item number of the original, operator's name, and date of capture. The accession and item number, uniquely identifying every image, can be used to link an image with a record in a bibliographic database.

Frame Grabbers

Another method of image capture is the use of a frame grabber, a piece of analog video equipment that is less expensive than digital equipment. A frame grabber bridges the analog world of traditional video technology and the world of digital imaging. It allows the use of consumer video technology for initial image capture, with subsequent conversion to digital form. A frame grabber is attached to a video feed and stores and digitizes a frame from a video sequence in various ways. Analog (such as National Television System Committee or NTSC video signal) encoding of images is more compact than full-digital representation, so a user often finds images stored on analog videodisc and transmitted using cable television technology. A frame grabber can provide the interface between this group of technologies and a digital environment of high-performance desktop workstations. One popular use of frame grabbers permits a workstation user to watch television programs pulled from a cable television network on half a split screen window while the user does programming or sends electronic mail on the other half. In this situation, a frame grabber digitizes images from the television signal many times per second and passes them to the workstation window manager for display.

Scanning Workstation Hardware

With rare exception, a digital scanner or camera has to have an associated scanning workstation, typically a PC or Mac. If the volume is low, the scanning workstation is the unit used for editing the images. If the volume is high, the PC or Mac is used for the capture software and the storage of the images before they are sent to the editing workstation.

At a minimum, a scanning workstation should be a 450 MHz or higher PC or Mac with 96 MB of memory, 8Mb video RAM, 17-inch monitor, three expansion slots for peripherals, 9 Gb hard drive with expansion slot for an additional hard drive, high-density removable drive (usually a Zip drive) for backup and for transfer of the

images to the editing workstation, or network interface card (NIC) for uploading to a server.

Image Capture Training

The Northeast Document Conservation Center hosts a "School for Scanning" several times a year for digital project managers. Information about the school can be found at ***www.nedoc.org.***