OVERVIEW

The common dictionary definition of an image is a reproduction or a likeness, but most dictionaries do not define imaging, digital imaging, imaging systems, imaging technology, document imaging, or electronic imaging.

Definition

For the purposes of this report, digital imaging, imaging systems, imaging technology, and so on are defined as a group of technologies that combine to allow images to be captured, edited, stored electronically, organized, transmitted, and accessed. The images may be pages of printed materials, photographs, slides, motion video, or any other source document or original.

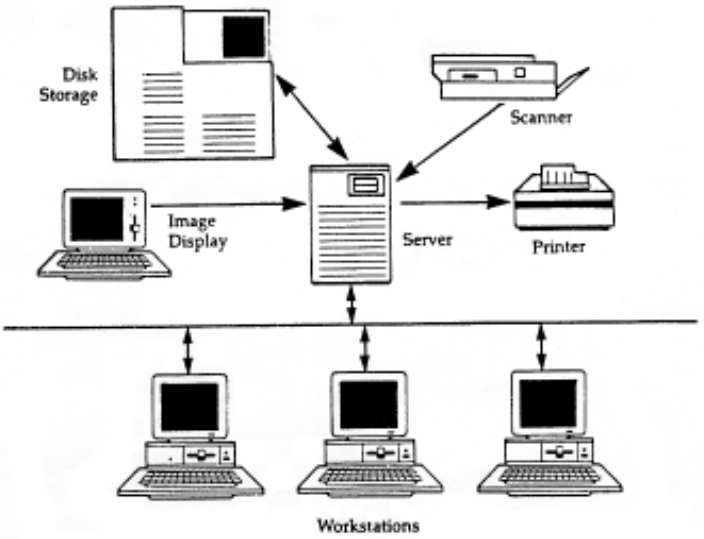
Distinguishing Digital Imaging From Graphics and Full-text

Digital imaging differs from graphics in that the former is concerned with the capture of that which can be seen by the eye and storing it digitally for subsequent access or retrieval; graphics is concerned with visual representations created from a database within a computer. Images require dramatically greater processing and storage requirements than graphics and, therefore, different hardware and software.

Images also must be differentiated from full-text. In digital imaging the image is digitized in its entirety so it is a visual replica of the original, and the latter is a character-by-character conversion to machine-readable form. The former allows only the retrieval of the image as a whole, and the latter allows searching of the text and retrieval of specific portions.

A Typical Imaging System

Although the technology can vary from application to application, the workings of a typical imaging system today are fairly straightforward. The following illustration represents the components and the process:



The source document is captured, edited, compressed to reduce the storage requirements, stored, organized (cataloged or indexed), transmitted over a network to the user who is seeking to access it.

History

As a business technology, imaging technology is relatively new. A spin-off of the sophisticated systems used by NASA to enhance and manipulate satellite images, it was used mainly by scientists and engineers until the early 1980s. At that point, imaging systems for business began to appear. They were generally customized, standalone systems that handled narrowly defined tasks and appealed to a limited market.

Among the technological advances that made imaging possible and increasingly affordable for a broad market are:

- Technology that provides efficient, high-resolution capture of images
- Improvements in processing power and memory of computers to facilitate the editing of images
- Storage that allows the electronic storage of the large amounts of data required for imaging at low cost
- Data compression to further reduce storage requirements
- High-speed networks that allow for the large bandwidth and high data transfer rates required to transmit images
- High-resolution display monitors to provide high-quality images.

Applications

Imaging technology can be found in many industries and institutions, especially the banking, credit card, and insurance industries in which the control of large amounts of paperwork is critical. Within these arenas imaging is displacing both hard-copy files and microform. Although microform is cost-effective and has a long life—assuming an appropriate, controlled environment—it also has disadvantages. Primary among these are the requirement for a specialized (and not popular among users) microform reader, and the fact that access to the material is limited to where the film or fiche and reader are located.

The medical community as well has been a leader in using imaging technologies. Picture Archive and Communications Systems (PACS) are rapidly replacing film for the capture, storage, and transmission of radiology images within hospitals. Other applications, including bedside instrumentation systems that capture patient chart information and combine it with x-rays and other medical records to provide a comprehensive clinical information system, are also becoming common.

For 15 years, libraries and archives have been using electronic imaging for the conservation and preservation of vulnerable documents and for the provision of access to source documents in a medium that is more efficient than microform. However, only in the past two years has the number of libraries and archives undertaking imaging gone higher than 200.

Imaging Versus Microfilming

The cost of image capture is greater than the cost of microfilming, but imaging

technology provides many options for editing of the images, such as the removal of stains or various adjustments to the image, whereas microfilming is not flexible. It is the enhancing of the images that represents most of cost difference between microfilming and imaging.

Stored images can be accessed by several users at the same time, including remote users. Images can be retrieved remotely if the transmission medium can handle the capacity requirements. Although images cannot be searched on individual words or phrases, re-editing a set of image records is always possible.

Frames from a microfilm can be processed to create digital images using special capture devices.

ADVANTAGES OF IMAGING SYSTEMS

In summary, imaging systems have several advantages, the most important of which are the ability to improved readability of a source document, dense storage capacity, speed of access, and the accommodation of multiple simultaneous users regardless of where they are.

Types of Images

There are types of images: thumbnail, service, and archival. A thumbnail image is a small image typically created to allow users to judge whether they wish to take the time to access a higher-quality image. A service image is designed to convey information that meets the needs of most users, so it has greater resolution than a thumbnail. It usually is compressed to save storage space and reduce bandwidth requirements for transmission. It takes longer to retrieve than a thumbnail. An archival image usually is an uncompressed image free of the artifacts resulting from compression. These are the highest quality images, but they require more storage and are slower to retrieve than either thumbnail or service images.

Capture

Devices are now available for capturing images at the rate of one every few seconds at resolutions ranging from 200 dpi (dots per inch) to 20,000 dpi. The former is equal to that output by a fax machine; the latter equivalent to a high-quality medical illustration. Capture devices range from inexpensive ($400) desktop models that capture a source document in as many as 30 seconds at low resolution to expensive (up to $40,000), production-oriented machines that scan in as few as three seconds or at resolutions of up to 20,000 dpi.

The choice of a capture device is the most important decision a library or archive makes because it determines the maximum quality that can be achieved regardless of the other components of the system. A source document captured at a high resolution can be displayed at a lower one; but one captured at a low resolution cannot be displayed at a higher one.

Editing

Frequently source documents, photographs, maps, and artwork are in poor condition. Besides fragility, they may have lost contrast, color definition, and clarity. In addition to physical remedies for the source documents, electronic remedies are available.

A high-end PC with a high-resolution monitor (1,600 by 1,280 dpi) and the appropriate software can be used to edit images to improve their readability. The images can be enhanced to repair damage, remove stains, or bring out hard-to-read

areas. Although most businesses do not seek to enhance images because they are scanning standard business documents and correspondence, many libraries and archives are addressing poor-quality originals. The software used to edit images in libraries and archives is usually Adobe PhotoShop or Corel Photo-Paint, both developed for professional photographers.

The editing can be limited to de-skewing (straightening the image) and cropping (removing unwanted margins), therefore, doing nothing to change the appearance of the source document; or it can include improving brightness, contrast, despeckling (removing discolorations or blemishes) and other means for improving the quality of an image.

Storage and Compression

All images require a considerable amount of storage capacity. Thumbnail images, which are stamp-sized, are usually 10 to 35 Kb each and are quickly retrieved and displayed. Service images, which typically are compressed using a format called JPEG, require up to 300 Kb each and can be retrieved in a matter of seconds. Archival images—the highest quality images—require large amounts of storage (several Mb each), so they are often stored on CD-ROM or DVD, rather than magnetic disk drives.

A single 18 GB magnetic disk drive can accommodate the equivalent of three to 7.5 million standard-size pages or 180 colored photographs of works of art. That ability is due not only to the storage capacity of the disk but also to sophisticated compression techniques that essentially reduce the storage requirements of a digitized image. This is important because images require a lot more memory than regular text files. Without compression, a large engineering drawing, for instance, could constitute 8 Mb of information and a high-resolution image of a work of art could require up to 100 Mb.

Compression is achieved by using the computer to remove extraneous information and condense redundant patterns of pixels—the basic elements of an image—into a kind of coded electronic shorthand. In this way, images can be compressed 20 to 30 times, reducing a large engineering drawing from 8 Mb to under 300 Kb.

Organizing

Most imaging systems include an indexing feature so images can be identified. In business that usually is a purchase order number, policy number, customer number, or date and a name or descriptor. Libraries and archives typically use a classification number or file number and a subject. However, many libraries chose to create full-MARC bibliographic records on their automated library systems and to link these to the images. There are limitations to MARC as it was not designed to handle some of the information common to images captured from source documents. There is work underway to develop new cataloging practices for all forms of information in digital form.

The Dublin Core metadata standard will be available at www.niso.org.

Two alternatives to MARC are the Dublin Core and EAD (Encoded Archival Description). Both are more flexible, but neither is a standard or widely adopted, so creating union databases in the manner libraries have done using MARC bibliographic records is difficult. Dublin Core is expected to become a NISO (National Information Standards Organization) standard in 2001.

Transmission

The transmission of images often challenges the capabilities of current networks. Although images are now being transmitted over networks via standard file proto-

cols such as TCP/IP and the Z39.50 computer-to-computer protocol, many networks lack the great bandwidth required to transmit images in a reasonable period of time. Transferring a single archival image over a 56 Kbit (56,000 bits per second) line could require up to 18 minutes. Few users would be prepared to wait for a multi-image file transmitted at that rate. On the other hand, a thumbnail would require less than a second and a service copy as little as five seconds. That is why access to archival images tends to be limited to access via a broadband LAN, rather than the Internet.

LAN is the abbreviation for local area network.

At a minimum, an organization that pursues an imaging program should design its LAN to support a bandwidth of 10 Mbps, and ideally 100 Mbps.

Access

Although a library or archive has no control over the choice of monitor by a remote user, it should plan on having high-quality monitors available for patrons within its facility(ies). At a minimum, they should be 19-inch color monitors that support at least 1,024-by-768-line resolution. Many older PC monitors display images at much lower resolution, so the resolution achieved with most scanners is lost.

Printing options also vary, from desktop laser printers for monochrome with 300 to 600 dots per inch (dpi) to the more expensive gray scale and color laser printers.

Issues in Imaging

The most pressing issues for imaging as a whole lie in the area of standards. As indicated above, data compression techniques are critical for reducing the storage required for images and standard compression algorithms are essential. Recognizing this, the Joint Photographic Experts Group (JPEG) and Motion Picture Experts Group (MPEG) have adopted standards for still and motion images. However, there are competing standards.

JPEG is the abbreviation for Joint Photographic Experts Group.

MPEG is the abbreviation for Motion Picture Experts Group.

Standards for the transfer of images are also required. Today, the Tagged Image File Format (TIFF) is commonly used, especially within the PC marketplace, but TIFF is only a *de facto* standard. PDF, also a *de facto* standard, is increasingly being used not only on the Web, where it originated, but also on in-house systems.

TIFF is the abbreviation for tagged image file format.

Color value standards for display devices are critical for color imaging to be viable. These are still in development.

Computer systems and network architectures also need to evolve to support open access to stored images. The current trend toward more distributed forms of computing, especially client-server architecture, is likely to combine with advances in high-speed data transmission to allow images to be integrated across applications.

Cost also remains an issue. Although scanning devices have made great technical, price, and performance strides in the last five years, the time and costs associated with creating an image database can still be high. This is especially true for databases of large amounts of nontextual data.

As of 2000, the only areas in which cost-benefits analyses have established the value of imaging have been banking, finance, insurance, and other applications involving millions of pages of records. Nevertheless, smaller systems are becoming more popular and prices are coming down. Also, more vendors of general purpose systems—including those marketing automated library systems—are beginning to enhance their functionality with imaging.

As a rule-of-thumb, a library or archive should expect the typical imaging project that includes a mix of source documents (text and photographs of various sizes and conditions) to cost an average of $6 per image. If microfilm or fiche of the source

documents is available, capturing the images digitally can cost as little as $0.20 per image; however, considerable additional expense is incurred if the images are enhanced after capture.

Determining whether images are protected by copyright and, if so, establishing the appropriate level of compensation to the copyright holder also must be addressed.

Availability of Turnkey Systems

Many specialized imaging systems are on the market, including products for banks, insurance companies, and hospitals. Systems designed specifically for libraries and archives are just now beginning to appear.

This report focuses on turnkey systems—those offered by a vendor that brings together hardware, software, installation, training, and ongoing support—which are potentially suitable for use by libraries and archives.

Types of Document Image Management Systems

Because different organizations have different information storage and retrieval requirements, there are different types of document image management systems. They can be categorized as follows:

- Desktop systems are designed for single-user applications. They handle a small number of images—usually no more than a few thousand. They are usually configured on a high-end PC. Although prices begin at under $10,000, they usually are not good choices because only a single user can be accommodated at any one time.
- Departmental systems, which are configured on an entry-level or mid-range server, handle five to 100 concurrent users and hundreds-of-thousands of images. The systems usually support access by clients via a LAN, a WAN, or the Internet. Prices begin at $25,000 and may go higher than $250,000.
- Enterprise-wide systems are designed to accommodate all of a large organization's documents and users. These can accommodate many hundreds of concurrent users and millions of images. The systems are typically supermicro-, mini- or mainframe-based. Image workstations may be directly connected or may come through a LAN. Multiple sites may be connected in a wide area network (WAN). Internet access is also possible. Prices begin at more than $250,000.

Selecting a Document Image Management System

Before evaluating imaging systems and vendors, having a thorough understanding of your application is critical. Document image management applications can be categorized into two types:

- *Workflow or transaction processing systems.* These systems focus on the processing of documents and aim at automating that processing. Examples include purchase orders, invoices, credit card charges, and insurance policies.

- *Storage and retrieval systems.* These systems are designed specifically to store and retrieve large numbers of documents. These documents can be a variety of types and formats.

This study focuses on the latter because that is the primary application in libraries and archives. Unfortunately, most of the literature emphasizes the former because that has been the priority for the users in the financial and insurance sectors and government, the major users of the technology.

Next you should determine whether the goal of the imaging project is preservation or improved access. The format emphasizes capturing images and storing them with little or no alteration; the latter emphasizes capturing the images and enhancing them to facilitate readability even at the expense of altering the appearance.

The size of the imaging system is substantially greater when the goal is access because the priority is usually to capture images of source documents that are likely to be in demand.

The Importance of Planning

Investment in imaging technology is not merely a matter of numbers. As with any introduction of advanced technology, there is considerable planning required and probably an extended implementation schedule. Document imaging represents a new process rather than simply a new tool. Although technical expertise is needed to design and install these systems, an understanding of the organization's goals and personnel issues is just as crucial to success.