SCANNERS FOR DIGITIZATION: AN OVERVIEW

Md. Nurul Alam*

* Library and Information Services,
Physical Research Laboratory,
Ahmedabad, Gujarat, India.

QR Code

ABSTRACT: -

Digitization of the resources of the libraries and the information centres has become very essential for preservation as well as better access of the resources. Selecting appropriate equipment for digitization projects based on the items is one of the important plans. Here an overview has been given about the basics of scanner and some of the important technical issues that are needed to carry out a digitization project. Emphasis is given on the factors of image quality at the point of digital capture and issues which impact selection of scanner types. The basics of what digital images are, including discussion on resolution, bit depth, and dynamic range, and how these concepts relate to creating good digital reproductions of various types of items such as text, manuscripts, and photographs. The purpose of this document is to offer guidance and to provide recommendations to LIS institutions which are planning to introduce systematic scanning of documents or are involved in more ambitious digitization projects.

KEYWORDS: Scanner, Digitization, Library and Information Science.

INTRODUCTION

Over the past few years, library digitization has had a major impact on the library science field. From the Library of Congress to academic libraries, public libraries, special and organizational libraries, digitization is today part of the work of many libraries in India and across the world. With more and more libraries digitizing their collections, policy concerns and technology problems surrounding digitization are becoming increasingly paramount (Liu, 2004). Digitization is one initiative that has changed the entire concept of libraries and the ways in which scholars, students, and users in general access and use scholarly information. Digitisation is the process of transforming the information from documents such as a printed book, picture or video into bits. Bits are the fundamental units of information in computer systems. Turning information into these binary digits is called digitisation. Digital forms of documents are

created through a variety of technologies. For creating digital collections in libraries, one needs to transform the existing print materials into digital collections. The processes and technologies connected with digitisation are quite complex in nature **(Rieger, 2008).**

Digitization promises to revolutionize the way libraries access, store, disseminate, and preserve information. Best practice, and its associated standards and technologies differ from institution to institution. Ultimate purpose of many digital projects is to preservation of historical resources and to provide greater access to collections, in order to contribute to public awareness and education, and to further research endeavors. The question of whether or not they will be better preserved in digital format is debatable **(Liu, 2004).**

PURPOSE

The purpose of this document is to offer guidance and to provide minimum recommendations that are planning to introduce systematic scanning of documents or are involved in more ambitious digitization projects. The recommendations in this document are purposely broad enough to apply to a variety of context. This document addresses standard formats of text and printed documents to be scanned These guidelines have been developed in order to increase the interoperability and accessibility of digital collections across the world through the use of widely accepted standards and formats, ensure a consistent, high level of image quality and decrease the likelihood of rescanning in the future by promoting best practices for conversion of materials into digital format and the long-term preservation of these digital resources.

CHALLENGES

When starting a digitization project, there are two items we cannot live without: a computer and a scanner. Acquiring the right hardware is one of the most important tasks in building a digitization project. Purchasing quality products that are also cost effective might seem like a big challenge, but with a little time and research can prove rewarding. One of the first decisions in creating the infrastructure of a digitization project is which scanner is to be used. From the literature it is difficult to find out any technical information about digitization project like what type of scanners have been used? What does digitization equipment do? What do digitization standard mean? What is the scanning software? What resolution was used? What file format was used?

SCANNING AND SCANNERS

When starting a digitization project, there are two items one cannot live without: a computer and a scanner. Acquiring the right hardware is one of the most important tasks in building a digitization project. Purchasing quality products that are also cost effective might seem like a big challenge, but with a little time and research can

prove rewarding (IFLA, 2002). One of the first decisions in creating the infrastructure of the digitization project is which scanner would be used for the scanning purpose.

There are many organizations that are interested in creating digital images of their collection like stamps, covers, and other philatelic material. There are many uses for these images, including increased efficiency of the process of working with documents, Quick and easy access to documents and information from anywhere, sharing of documents throughout the whole organization, saving time and cost of manual data processing, secure storage of documents with instant access and preventing the loss of documents (Business News Daily, 2016). A scanner is a device that converts images to a digital file .The process of creating a digital image from a paper document is called scanning.

A variety of scanning devices exist for the automatic capture of spatial data. While several different technical approaches exist in scanning technology, all have the advantage of being able to capture spatial features from a map at a rapid rate of speed. Scanners are generally expensive to acquire and operate. As well, most scanning devices have limitations with respect to the capture of selected features, e.g. text and symbol recognition. Experience has shown that most scanned data requires a substantial amount of manual editing to create a clean data layer. Given these basic constraints some other practical limitations of scanners should be identified (Buckey,(n.d))

FACTORS CHOOSING A SCANNER

There are some factors to be considered while choosing any kind of scanner, these are mentioned below:

- **Maximum paper size (Scan area).** The largest size paper that can be passed through the document scanner. Usually A4 or A3, unless you're looking at specialist document scanners.

- **Daily throughput or duty cycle.** The number of pages that can be scanned each day. This will range from a couple of hundred to many thousands, depending on which category the document scanner falls into.

- **Scanning side.** Whether the scanner can scan one or both sides of the paper automatically. Simplex means single sided. Duplex is double sided. If you will be regularly scanning double sided documents, opt for a duplex model.

- **Resolution.** The level of detail at which your document scanner operates. This is measured in dots per inch (DPI). Although lots of factors affect the overall quality of scans, in general a higher the DPI is better.

Ignore 'interpolated' DPI figures – always look for the 'optical' figure.

- **Colour mode.** Can the document scanner handle full colour, black and white or greyscale? Most scanners can scan in full colour.
- **Pages per minute (PPM) and images per minute (IPM).** The number of pages the document scanner can scan in a minute. 'Images per minute' applies to duplex scanners, because the front and back are scanned simultaneously, giving double the number of single pages. Speeds are usually quoted for a particular type of document, like A4 black and white scanned at a certain resolution.
- **Automatic document feeder (ADF) capacity.** The number of pages you can load in for scanning in one batch. With a large capacity, you can set the scanner up, then go do something else while it scans.
- **Flatbed scanning.** Flatbed document scanners can scan documents that cannot be loaded into an automatic feeder. These items include bound documents and books. Some scanners have both an automatic document feeder and flatbed facility.

- **Interface or connectivity.** Make sure the computer to which you intend to connect your document scanner has the right connections. USB has become the common standard, but you may also see SCSI or FireWire.

- **Drivers and compatibility.** Drivers are files that sit on your computer and enable it to 'talk' to your document scanner. Make sure your chosen document scanner will work with your computer and its operating system. And if you're planning to use your document scanner with software that doesn't come with the scanner, check the drivers will work with it.

- **Bundled software.** Most document scanners come with software. This can range from a trial copy of basic scanning software to a full commercial product worth hundreds of pounds. Purchasing a document scanner with a comprehensive software bundle can reduce the overall cost of your scanning solution – as long as you need the software.

KINDS OF SCANNER:

Flatbed scanners

Flatbed scanners, also called desktop scanners, are the most versatile and commonly used scanners. A flatbed scanner is made up of a glass pane and a moving optical CIS or CCD array. The pane is illuminated with the help of bright light planted underneath it. The image which is to be scanned is then placed on the glass pane. The sensor and the source of light move across the glass pane to scan

the document and produce its digital copy. The most common scanners are flatbed scanners. These are a type of reflective scanner that commonly sits flat on a desk. Flatbed scanners adapt scanning pieces of paper, objects, photo prints, and other opaque items. Examples of flatbed scanners that are in the market are: Canon - CanoScan 9000F and CanoScan 5600F; Epson - Perfection V750; HP - ScanJet N9120 and Fijitsu – fi-6230Z (Cornell University Library, 2003; ebay,2014)

Overhead scanners

A specialized variant of the flatbed scanner is the overhead book scanner, in which the scanner's light source, sensor array and optics are moved to an overhead arm assembly under which a bound volume can be placed face up for scanning (Cornell University Library, 2003;)

Sheet-fed Scanners

A sheet-fed scanner looks a lot like a small portable printer. Sheet fed scanners are fundamentally the same as flat-bed scanners, except that instead of laying the document to be scanner flat on the imaging surface, the document that is supposed to be scanned is fed into the horizontal or vertical slot provided in the scanner By their name, it is clear that only individual sheets of paper are fed into these scanners, so they do not offer the flexibility of flat-bed scanners, but they do offer greater performance, since sheets feed in automatically. The vital components of sheetfed scanner are the sheet-feeder, scanning module and calibration sheet. Such scanners are most often used to scan single page documents (Cornell University Library, 2003; ebay,2014)

Drum Scanners

Drum Scanners, the first type of image scanners ever invented, capture images using photomultiplier tubes (PMT), which is a type of vacuum tube that is highly sensitive to light which moves back and forth along a single axis. The image to be scanned is mounted on an acrylic cylinder which rotates at high speed in front of precision optics that deliver the image information to the array of PMTs. Modern drum scanners use 3 PMTs, which read red, green and blue.

The use of drum scanners has declined significantly as flatbed scanners based on charge-coupled devices have dropped in price; however, drum scanners are still used for certain high-end applications, such as museum-quality archiving of photographs, desktop publishing, and print production of books and magazines. They are very expensive. Even though few companies continue to produce drum scanners, they are still in demand due to their ability to produce high resolution scans. The obvious disadvantage of drum scanners is that they aren’t as portable and easy to use as flatbed scanners and take a long time to scan an image. The biggest disadvantage though is price (Cornell University Library, 2003)

Handheld Scanners

Handheld scanners use the same basic technology as a flatbed scanner .A handheld scanner is a small manual scanning device which is moved over the object that need to be scanned. For instance, if a document needs to be scanned, the handheld scanner has to be dragged over the document. Hand scanners are useful for their portability and low price (often one-third to a quarter of the cost of a flatbed scanner). Hand scanners generally plug into a computer's printing port, as opposed to a SCSI port, allowing them to be easily shared from workstation to workstation. Many people find them ideal for use with a notebook or laptop. Unfortunately, hand scanners are less accurate than flatbeds because they have weaker light sources and often produce uneven scans - courtesy of the unsteadiness of our hand or the surface we are standing on. Using a handheld scanner can prove to be a cumbersome task, as the hand needs to be steady all the time. Slight movement of the hand can lead to distortion of the image. One of the most utilized handheld scanners is the barcode scanner, typically used in shopping stores to valuate goods (Cornell University Library, 2003; ebay,2014)**.**

Photo Scanner

Photo scanners are mostly used to scan photographs. Photo scanners have evolved into high-resolution image processors capable of scanning photo negatives and producing high quality photo enlargements. High resolution and color depth are the most vital requirements for scanning photographs, and photo scanner provides the same. If the motive of buying a scanner is to digitize film negatives and slides, then the photo scanner is the best option. They are specially designed to work on slides and negatives. The in-built software in some photo scanners can also help in cleaning old photographs. Examples of photo scanners that are in the market are Canon CanoScan, HP Scanjet, Epson Perfection (Cornell University Library, 2003; ebay,2014)

Film Scanners

Film scanners are specialized transmissive scanners made to scan film strips and mounted slides (negatives and positives). Film scanners have optics and electronics specifically catered to scanning film. Because of this specialization, film scanners achieve better results when scanning film than flatbed scanners. A film scanner is utilized to scan photographic films directly into a computer. The photographer has direct control over certain aspects, such as cropping, ratio of original image on the film, etc. Some film scanners have specialized software through which it is possible to minimize scratches and improve color quality. Low-end film scanners most often accept 35 mm film strips while the high-end scanners have interchangeable film loaders which can accept 35 mm strips or 120 mm ones, or individual slides. Some film scanners include feeder attachments that can make doing many scans easier and faster. Some scanners, such as

the Nikon 5000 ED for smaller format films and Nikon 9000 ED medium format film. (Cornell University Library, 2003)

Portable Scanners:

Being small in size, portable scanners can be easily carried with oneself anywhere. Some of these are as small as our PDAs, hence, can be easily carried in the pockets. They are of great help when it comes to text document scanning. Their drawback though, is their limitation as far as resolution is concerned. They cannot be used for scanning photographs or other such applications which require high resolution scanning (ebay,2014; Buzzl, 2015).

Book scanners/planetary scanners

Books, newspapers and journals can be easily damaged in standard flatbed scanners. The book's binding or the fragile nature of the document may prevent the object from being held flat on the glass. This can result in uneven illumination and lack of sharpness, heavy books and brittle pages can also be damaged when placed face down in a scanner. The typical book scanner (also known as a planetary scanner) uses one or two high resolution digital cameras with two or more lights to provide even illumination. By using a camera to capture the image, the object does not have to come into direct contact with the capture device. Some book scanners use a single camera above the object, the book's spine and cover is supported in a frame and the photographed page is held parallel to the camera. A few book scanners use a V-shaped cradle and two cameras, one camera captures the right side, the other the left. The V-shaped cradle places a minimum of strain on the book's binding and pages, the cameras are positioned so that they are parallel to the page they are capturing

Robotic scanner

A cutting-edge robotic scanner – the first in the UK and only the second in the world to be installed in a research library – is being used in an exciting initiative to create a vast digital library from original bound and printed historical documents. The University of Southampton is using the unique precision-built equipment to scan rare parliamentary documents as part of a project that aims to put 300 years of history online. Southampton is leading a consortium of researchers and academic libraries to digitise all surviving 18th century parliamentary papers and bills. These will then be available in a comprehensive web archive for students of history all over the world to access directly from their home or office computers (University of Southampton, 2004)

Scanner type	Flatbed	Sheetfeed	Drum	Slide/Film	Microfilm	Digital Camera
Sensor Technology*	CCD/CIS	CCD	PMT	CCD	CCD	CCD/CMOS
Media Types**	R, T	R	R, T	T	T	R, T
Tonal Range***	B/W, G, C	B/W, G	G, C	G, C	B/W, G, C	G, C
Typical Size Limitations (inches)	11x17 8.5x14 w/ADH**** 8.5x11 low-end	8.5x14 tabletop 24x36 upright	12x17	35mm, 4x5	35mm. Some handle 16mm & 105mm, fiche	None
Resolution Range (dpi)	300-1200	200-400	1200-8000	1000-5000	5000-8500	Pixel array: 1600x1200—12000x10000
Price Range (US$)	50-50,000+	5,000-100,000+	10,000-100,000+	500-30,000	40,000-250,000	200-40,000
Typical Use	single leaf, bound volumes (overhead type)	uniform bus. docs., oversize	pre-press	orig. slides, intermediates	film, fiche, aperture cards	bound volumes, 3-D, fragile, oversize

Table: Comparison of Scanners (Technical Attributes/ Parameter)

* CCD=Charge Coupled Devices; CIS=Contact Image Sensor; PMT=Photomultiplier Tube; CMOS=Complementary Metal Oxide Semiconductor ** R=Reflective; T=Transparent *** B/W=Black and White; G=Grayscale; C=Color
****ADH=Automatic Document Handler

(Cornel University Library, 2003))

Scanner Sensors

Scanner image sensors help determine the quality of scans. The sensors convert the light reflected onto the object being scanned into a clear digital image. Scanners typically use one of three types of image sensors: charged-coupled device, contact image sensor, or photomultiplier tubes. Charge-coupled Device (CCD) - CCD scanners work in conjunction with a bright light that illuminates the image. Then, the CCD uses its rows of pixels to capture the image. CCD scanners usually produce better image quality than CIS scanners. Contact Image Sensor (CIS) - CIS scanners, the image sensors lie directly under the document and picks up the reflected light right from the document. CIS scanners tend to be more compact and durable than CCD scanners. Photomultiplier Tubes (PMT) - In drum scanners, three photomultiplier tubes are often used to capture red, green, and blue light separately. PMTs are more sensitive than CCDs and CISs, so they can create more detailed scans ((PARADIGM, (n.d))

CIS vs CCD

Sensor	Advantages	Disadvantages
CIS (Contact Image Sensor)	Less cost High reliability High productivity More compact No stitching needed Higher Optical resolution High MTF (Contrast) sharper lines No lens distortion	Sensitive to focus depth Lower Signal/noise ratio due to light source
CCD (Focused Image Sensor)	High Signal/Noise ratio due to light source Relative insensitive to focus depth	High cost More complex and fragile technology At least twice the size CIS technology Requires frequent stitching &

		calibration Lower optical resolutions Lower MTF (contrast) blur red lines Inherit Lens distortion

(PARADIGM, (n.d))

Summary

To sum up the world of scanners, it needs to be mentioned that at its very core, digital imaging is governed by important performance parameters like pixels, grayscale, bit-depth, resolution, color, and file size etc. These parameters are conceptually different for image acquisition equipment like scanner as compared to image output and viewing devices. Set against this theme, Scanner as well as scanning process should recognize few basic concepts in digitization project. The present methodologies for digitization in India, involve extensive handling of digital images. Scanners have become one of the most common instruments for taking digital images and generally none of the digitisation projects can manage without a scanner as their input devices. In this context, this paper can provide an overview to the beginners who plan to embark upon a digitization project.

References

REFERENCES

- Buckey, D.J. Bigs introduction to GIS. http://bgis.sanbi.org/GIS-primer/page_25.htm
- Business News Daily. (2016). Choosing a Digital Copier for Your Business. https://www.google.co.in/url?sa=t&rct=j&q=&esrc=s&source=web&cd=1&cad=rja&uact=8&ved=0ahUKEwiaxuPR-KjLAhVTbY4KHXk7BdUQFggbMAA&url=http%3A%2F%2Fwww.businessnewsdaily.com%2F7898-choosing-a-digital-copier.html&usg=AFQjCNFXBTyTnj-ezTy2SdAFukNqXZdo-Q&bvm=bv.116274245,d.c2E
- Buzzle.(2015). Different types of Scanner. http://www.buzzle.com/articles/different-types-of-scanners.html
- Cornel University Library. (2003). Moving Theory into Practice: Digital Imaging Tutorial.https://www.library.cornell.edu/preservation/tutorial/technical/technicalB-03.html
- Ebay.(2014). Scanners Buying Guide. http://www.ebay.com/gds/Scanners-Buying-Guide-/10000000177628791/g.html
- IFLA. (2002). Guidelines for Digitization Projects for collections and holdings in the public domain, particularly those held by libraries and archives.http://www.ifla.org/files/assets/preservation-and-conservation/publications/digitization-projects-guidelines.pdf

- Liu,Y.Q. (2004). Best practices, standards and techniques for digitizing library materials: a snapshot of library digitization practices in the USA, Online Information Review, 28(5), pp.338 – 345
- PARADIGM. CIS vs. CCD Scanning Technologies http://www.scantopia.com/Pdf/Graphtec/CISvsCCDver_final_Version.pdf
- The IT Donut. (2016). How to choose a document scanner. http://www.itdonut.co.uk/it/computer-hardware/computer-printers-and-scanners/how-to-choose-a-document-scanner
- University of Southampton. (2004). Robotic scanner preserves the pages of history. http://www.southampton.ac.uk/news/2004/11/scanner-preserves-pages-of-history.page