

How to Evaluate a Book Scanner

By Derek Jenkins

Introduction:

The focus of this paper is to provide an introduction to the complex issues of evaluating and choosing a book scanner for use in the library and archival work place. There is no right or wrong answer, but being educated allows you to make a more informed decision. This report is primarily aimed for the non-technical user, but does introduce some technical issues that I will cover as simply as I can. For more detailed and technical information on evaluating scanners, experts like Don Williams (ex-Kodak), Steve Puglia (NARA) and the archival staff at LOC can provide additional resources.

The issue:

Choosing a scanner to scan books and documents can be a daunting task. The industry has systems available from a few thousand dollars to over a hundred thousand dollars. All of these can scan, but is the way that they scan going to meet your project objectives? Is the scanner within your budget? Is the scanner even within your staff's technical capability? In many ways the budget drives the selection, but buying a sports car to take a large family to Yellowstone when the number of people was better suited to a minivan will make the task very difficult.

Here is a list of my evaluation criteria in order of decreased importance. However, the points are interrelated so selection is commonly a compromise among all the points.

Quality: This is the main topic in this paper and, in my opinion, the most important factor when buying a scanner.

Reliability: If the scanner is always failing and needs constant maintenance then it will not be available for the task it was intended.

Performance: If the time it takes from one scan to the next is too long then scanning more than a few pages at a time is not practical.

Functionality: Knowing what you will be using the scanner for will help define the functionality required in the system.

Cost: Cost is a trade off between quality, reliability, performance, and functionality.

How to Evaluate a Book Scanner

By Derek Jenkins

Background:

1. At the basic level, all scanners capture an image in Grey Scale or Color. The associated hardware or software defines what image is finally delivered you.
2. All scanners capture an uncompressed data stream that creates an image when viewed in two dimensions.
3. Scanners DO NOT scan in PDF, JPG or TIFF. These are the output formats from the software that is running the scanner.
4. You can only evaluate the quality of a scanner by looking at an uncompressed image. You can not evaluate the scanning quality by looking at an image that has been compressed as the image is degraded during compression.
5. Images from a "SIMILAR" model system should never be used to evaluate a different scanner model. Different models use different cameras, electronics, lighting and scanning methods. This is why it is sold as a different model.

Hardware and Image Formats:

Images, when scanned, normally arrive in the memory of the PC in either grey scale (8 bit) or color (24 bit) depending on the camera being used. From that point on, any output images are a *derivative* of the original scanned image.

For instance, a bi-tonal image is made by converting an image from its original grey scale or color format to a 1 bit mono-chrome representation. This can be done in many ways, perhaps with a simple static thresholding or with a more sophisticated dynamic threshold, but at the end of the day it is still a software mathematical algorithm that produces the output. This algorithm has nothing to do with the quality of the scan that was originally captured by the scanner.

Once in the PC's memory, the captured image is then normally compressed for storage. The most common method for bi-tonal image storage is called CCITT-G4 compression. This 'compressed' image is then wrapped in a TIFF header that contains other basic image information like the image width and height. Hence the common term of TIFF-G4.

Some of the typical compression schemes for color and grey scale images are JPEG, JPEG2000 and JBIG. They take an image, compress its size with a mathematical algorithm and store it on disk. In most cases these formats achieve their savings in size by introducing a slight "loss" in the image quality. The higher the loss ratio or percentage the smaller the file size but at the expense of increased image degradation.

How to Evaluate a Book Scanner

By Derek Jenkins

All compressed image formats are images that have been mathematically manipulated and degraded to achieve a smaller storage foot print. Therefore, these images are not valid to be used for scanner quality evaluation. **The only true representation of the scanner's image quality is an uncompressed color or grey scale image that is typically stored in a RAW or TIFF format.**

Image Quality Evaluation:

For evaluation, it is important to consider scanning some of your proposed work. However, this still provides a very subjective and limited review. Later projects may have an entirely different type of material. Testing with one project in mind may not show inherent scanner weaknesses that could be an issue later on. A better method is to scan a known entity like a standard scanner target. A target simulates the basic components of all images to allow evaluation of a scanner's quality. (See *Applied Image at the end of this paper.*) These targets are available in many different layouts and most will do an excellent job of demonstrating the strengths and weaknesses of the scanner. Working with this known entity allows removal of the personal feelings about the documents and brings the evaluation to a more technical level.

Don Williams, an Image Quality and Calibration Consultant and formerly a research imaging scientist in the Imaging Science Division of Eastman Kodak has been working with NARA and the Library of Congress to produce an improved mathematical method to evaluate scanners. This process started with the SFT targets that have evolved into the SFT II targets and now a new version specifically for book scanners will soon be available. Considering the type of work that you are doing this may be little overkill, but it is the ultimate method for scanner evaluation.

Without going to extremes, scanning any of the commercially available targets will allow you to evaluate many of the basic functions of a machine. However, you must insure that the images being evaluated were produced on the exact model with the exact lens, lighting and camera that you are considering purchasing. As stated earlier, allowing samples from "something similar" will make the comparison totally invalid.

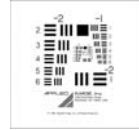
A scanner consists of multiple components like camera, interface boards, lens, lighting and transport that, when used together, produce an image. One of the evaluations of a scanning system is to view the sharpness or focus of an image before any enhancements. The finer the details viewable, the better the scanner. In optical terms this is called the Modular Transfer Function or MTF. Quantitatively measuring MTF takes a special program but a simple evaluation can be made by looking at certain parts of a scanned target.

How to Evaluate a Book Scanner

By Derek Jenkins

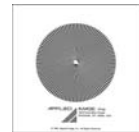
When evaluating images they should all be viewed on the same monitor at a 1:1 or 100% zoom. Other levels of zoom scale the image up or down and do not provide a true representation of the original scan.

Line Pair per Millimeter. These are read by looking for the highest numbered line pair on the target that you can distinctly see the black and white lines. The higher the number the better the image *at the resolution scanned*. Don't compare 200dpi on one machine with 300dpi from another.



Fan Wedge Target. This is a fan shaped series of lines that are thicker at the bottom and thinner at the top. This is used like the LPPM and you look for the highest number where you can see the distinct difference between the black and white lines. The higher quality scanner will have the higher line delineation. In addition to the simple numeric reading you can also compare the straightness of the target lines.

Pestrecov Star Sector Target. This is a circle of lines, thicker on the outside and thinner at the center. This is evaluated by looking to see how close to the center you can still see the distinct black & white lines. Again, the better scanner will show distinct lines closer to the center.



Text Size Targets. This is the alphabet in upper and lower case printed in ever smaller type fonts. The evaluation is performed by comparing which lines of text can be read. The best scanner will render the smallest fonts clear and legible. This is very important when scanning books. Footnotes, especially, may require a scanner that can scan smaller fonts than some scanners can provide.

Straight line Test. This is a line normally drawn diagonally on the big target. The very simple evaluation is to review how straight this line is. A significant wave distortion in the line indicates major issues with the scanner.

Color Separation Test. This is my favorite test because it tells so much about the scanner and is so easy to see. All one need do is to look at any part of a target with a solid black vertical or horizontal edge. This could be lines or text or anything that has a good transition from black to white. Look at the line at 100% and see if you can see any color shadow or halo. Then zoom to 200%, 300%, 400% or more looking to see if this color shadow is seen. Look on both sides of the line or character. Color halos at the default 100% reveal a low end scanner or a bad array. Average quality scanners will not show a color halo until 400% or so. The best scanners will not show any haloing at all.

How to Evaluate a Book Scanner

By Derek Jenkins

Full area Scan. In this test, the scanner's maximum area size is fully utilized. A target is placed in each corner of the area and one is also placed in the center of the scan area. Evaluate the images for optical distortion. The poorer quality the scanner the more differences will be viewed between each target's results.

Image target evaluation must be done at different resolutions. Some scanners like the i2s Digibook have a variable optical system. This means that the operator moves the camera and refocus the lenses to achieve different resolutions. Other scanners have a fixed optical resolution. This means that the camera and lens are fixed at a given resolution, like 300dpi. The image's resolution is scaled down (pixels removed) with software for lower resolutions like 200dpi. Higher resolutions than the native image from the scanner (like a 400dpi resolution from a 300dpi scanner) is scaled up. Scaling down can be an acceptable image as it is a derivative of the original but scaling up involves interpolation and adding pixels that were not in the original scan. This is not acceptable if precise image representation is required.

I have stayed away from true color evaluation because there are many more steps involved than just scanning a color target. If a scanner has the ability to generate and use an ICC profile then color variances can be corrected. I am not saying that scanning a GretagMacbeth® color target is not useful. I am saying that true color evaluation is only possible under very controlled circumstances with special calibration tools. This requires the scanner and monitor both be calibrated and an ICC profile generated. This step insures that the monitor is reproducing a perfect replica of the scanned original. You would only be able to confirm that the monitor is doing this perfect replication if you view the original document in a special calibrated light box. This detail is beyond what most people can practically do. Entities like NARA or OCLC that have goals for the Pristine Archival scanning buy the equipment and search out the expertise and training to do this type of "full" color evaluation. As long as the scanner produces decent eye discernable color and has the ability to use ICC profiles then this generally is all that you should need in most cases.

Reliability:

If the scanner is constantly failing and needs frequent maintenance then it will not be available for the task it was intended. These ongoing repair costs will effectively increase the purchase price of the scanner so what appeared a good buy suddenly becomes a burden to the operational budget.

Performance:

A manufacturer's documentation may state that it takes one (1) sec to scan an image. Check carefully what it means by that. There are several elements to the action of scanning. A scanner may take just 1 second to do the scan but it also

How to Evaluate a Book Scanner

By Derek Jenkins

may take additional time to set the exposure, correct the curvature and save the image to disk. The time that should be looked at is not the scan time but the full cycle time that takes all these steps into account. The cycle time is the time from the start of a scan to the time you can start the next scan, repeatedly. Depending on your goals (the number of materials you have to scan and the amount of time you have to complete a project), this cycle time could be too long and then scanning more than a few pages at a time is not practical.

Let us look at this from another angle. If I am scanning a book using a glass plate, how fast can the operator release the glass, turn the page, close the glass and then press the scan button? If that time is less than the scanner's save time then the operator will be waiting for the scanner. If it is more than the save time then the scanner is waiting for the operator.

Performance is not everything, but it is important. I would much prefer that the scanner is waiting on me to finish turning the page than me waiting for the scanner to finish saving. I have seen a scanner run at 12 scans per min or about 24 pages per min. This is about as fast as an operator can work for any extended time.

Functionality:

What ability do you want your scanner to have? Is this functionality a part of the scanner or part of the software tool that is driving the scanner? Why and when does this matter to you?

1. If the majority of your books and documents are 8.5x11 inches or 11x17 inches do you really need to have a scanner that will scan the occasional 22x34 inch map?
2. Do you need a book cradle or is most of your work flat?
3. Do you need a glass platen to hold the work?
4. Do you need a vacuum table?
5. Do you want the scanner to just scan and then you will post process all images later with other software like Photoshop®, or BookRestorer® or some other software program?
6. Do you want the scanner to deliver finished work at the expense of speed?
7. Do you want the scanner to be portable?
8. Are you scanning for pristine archival work?
9. Do you have operators that are capable of operating a complex optical system?

How to Evaluate a Book Scanner

By Derek Jenkins

Cost:

Many factors need to be considered. Comparing the cost of a system needs to be done on an apples to apples basis. Here are some questions to ask.

1. Does the scanner have a book cradle?
2. Is it mechanical or electric?
3. Does the system have a glass?
4. How does a color scanner scan grey scale? i.e. Does the scanner do grey scale at a native level or does it generate grey scale from a color scan?
5. Does the system have operator exchangeable cameras for higher (or lower) resolutions?
6. If the price seems too good to be true is this the end of a model's production life? Will it be replaced with an improved version soon?
7. Is the scanner's capability being exaggerated so that you will buy it? Should you actually be looking at that next model up to meet your goals even if it does cost more and may not be in your current budget?
8. Does the system need a PC or monitor and are they included?
9. Is freight included in the quote or is it billed later?
10. What is the warranty period? And how much does it cost to extend it?

Any purchase is a compromise of quality, performance, cost and functionality. What compromise are you prepared or need to take for the project that you are thinking of taking on?

Here are my suggestions:

Verify the information given in the advertising literature or provided by the salesperson. Check all the numbers in a real world setting like a trade show. Sit down at the scanner and use it like you would if you had purchased it. Be wary of statements like "the scanner was damaged in shipment" if the performance is not as advertised. Ask for multiple references and call them all. Try and find a user with the scanner that was not a reference since a company is not likely to give out the name of a site that had problems. Ask these references probing questions to explore and compare how you are planning to use the scanner. Ask the references about the scanner's reliability and the post sales support.

Ask the salesperson for samples to be done on a machine that is exactly like the one you are considering. If possible, be present when the samples are done. Maybe try to do them at a trade show. Also, ask for targets like the Kodak TL5003 "Old man target" to be scanned in both Color and grayscale and at different resolutions and save them as uncompressed TIFF images.

Ask for a scan of the GreTag color patch target at different resolutions and also save them as uncompressed TIFF. Lastly, scan an area the full



How to Evaluate a Book Scanner

By Derek Jenkins

size that can be handled by the system. Scan it multiple times moving a target to each corner. All this information will let you know the capabilities and limitations of the scanner.

Summary:

Scanner and image evaluation is a very, very broad topic. . I have tried to stay with the core points that can be looked at by most people. There is so much more that I have not even touched on like lighting, color temperatures and optical systems. I have also not covered the image distortion and its required mathematical correction that is inherent in some scanner designs.

This outline has been produced as a source of information for you. I have tried to be as generic as possible and not slanted in any direction. I am very willing to spend time and discuss this with you to help you produce good evaluation criteria.

How to Evaluate a Book Scanner

By Derek Jenkins

Company Background:

IImage Retrieval, Inc (IIRI), founded in 1990 by CEO and President, Derek Jenkins and has been involved in the microfilm digitizing and document scanning industry since its inception. Coming from a systems software and R+D background Derek allied his new company with the fledgling scanner manufacturer called SunRise® Imaging and participated in the development of a new microfilm scanner. Over the years SunRise® became the leader in the film scanning industry and IImage Retrieval was its largest and most experienced reseller and repair facility worldwide.

In 2002, IImage Retrieval severed ties with SunRise® and IIRI's principles formed another company called nextScan. They developed a new line of scanners and applications that set new quality and performance standards in the film scanning market.

Driven by IIRI's customer's requests, IImage Retrieval, Inc expanded its offerings to include the line of rare book scanners produced by i2s. The introduction in the US of the DigiBook™ line of archival scanners was heralded by the Ransom Center at the UT Austin digitizing their copy of the Gutenberg Bible for publication in their digital portal. Their Gutenberg edition is one of the better complete volumes in the US. Since then many other establishments have purchased scanners to digitize their special collections and archives including LOC, NARA, University of Pittsburgh, University of Virginia and the University of Florida at Gainesville. The University Of Florida has seven CopiBooks scanning a massive collection of news papers.

The i2s line of high resolution scanners were designed for the archival industry but also fill a very flexible roll scanning any large flat objects from museum displays, tapestries and pictures to maps drawing or newspapers.

Reference:

Applied Image, Inc www.AppliedImage.com Luke Hobson(585)482-0300
Image Calibration Targets

Don Williams

Image Quality and Calibration Consultant