Document Scanning Technology

Understanding the issues.
Making the right choices.

kodak.com/go/docimaging

A-5509 CAT No. 845 1304

Printed in U.S.A.

DOCUMENT
IMAGING
INNOVATION YOU CAN COUNT ON™

Kodak

Introduction

The purpose of this informational guide is to help buyers and users of production document scanners better understand the technology considerations influencing productivity and total equipment ownership costs.

Over the past decade, information-intensive industries, businesses and government agencies have made significant investments in document imaging technology to improve customer service, speed communications, process more information in less time, and improve the cost efficiencies of workflow. In many applications, enhancing customer, client or patient satisfaction has been a driving force in reengineering document management strategies that require an ability to move paper quickly through the capture process to produce high-quality, legible, digital images.

Evaluating Volume Requirements

Document scanner manufacturers design equipment that is best suited to meet a variety of page volumes, feature sets and price requirements. Common scanner market segmentation, introduced in 1999 by organizations such as InfoTrends Research Group, Inc., (Boston, MA) categorizes scanners by price and speed ranges. The segments are ***High-Volume Production***, ***Mid-Volume Production***, ***Low-Volume Production***, ***Departmental*** and ***Workgroup***. This guide will focus on production scanners.

High-Volume Production scanners achieve speeds of 60 pages per minute (120 images per minute) or greater. Some reach speeds of 200 pages per minute. Scanners in this category range in list price from $30,000 to $100,000 or greater. These speed demons typically handle 10,000-30,000 pages or more in a single workday.

Mid-Volume Production scanners have rated speeds of 42-85 pages per minute, and are designed to scan daily volumes of 5,000-10,000 pages. List prices in this classification are approximately $12,000 to $29,000.

Low-Volume Production scanners typically handle 36-50 pages per minute, and are designed for volumes of 500-4,000 documents per day. Units in this category carry list prices of $6,000 to $12,000.

Duplex and Simplex Scanners

Duplex scanners capture both sides of a page, in a single pass, to increase productivity and ensure that both sides are scanned together. Simplex scanners scan one side of a page at a time. To scan a double-sided page, each side must be scanned separately, then matched together electronically. This creates the potential for quality control problems – especially in batch environments. A good rule of thumb to consider: if your daily scan volumes contain 30 percent or more of two-sided documents, then it becomes more economical to consider a duplex scanner.

Color and Bitonal Scanners

In the past, color scanning options have generally been considered expensive and particularly lacking in production-level throughput. In just the past two years, color production scanners have been introduced featuring an attractive combination of price, image quality, and speed. The industry is beginning to offer more models of production color scanners with varying price and feature sets.

Businesses are rapidly adopting color production scanning for several significant reasons. Color scanning:

- Simplifies document capture
- Captures more image content, bringing out ***all*** the information
- Captures documents accurately the first time, with no manual adjustments
- Delivers a much higher-quality image at half the dpi resolution of a bitonal image
- Provides color image files that when compressed are not significantly larger than bitonal image files

Most color scanners also offer bitonal and grayscale output options. The result is that companies now have more options when choosing the most appropriate scanner for their application. Production color scanning is now just as valid a business choice as bitonal or grayscale scanning have been for years.

Scanning Speed Versus Throughput

Scanning speeds are typically quoted in either pages per minute (ppm) or images per minute (ipm), which are interchangeable values on a simplex scanner. However, on a duplex scanner, images per minute can be as much as double the speed of pages per minute. Be wary of estimating actual scanner throughput based solely on transport speed, which is typically expressed in inches per second, as this does not account for the gap between documents.

Scanning speeds are often published using the landscape scanning mode, which increases the apparent speed rating because of the shorter page length (8.5 inches for landscape

mode versus 11 inches for portrait mode). Many of the most popular capture software applications can automatically rotate images 90 degrees for viewing, so scanning speed can be improved by more than 20 percent.

In relation to scanning speed, throughput is more difficult to measure since it is affected by a wide variety of variables including paper type, size and quality; document preparation time; paper handling and feeder characteristics; scanner speed at a given resolution; the inter-document gap and the involvement of the operator.

To ensure that you choose the most appropriate scanner to meet the time and accuracy constraints of your business, it is a good idea to benchmark test your own "real-life" business documents. Comparing scanners through benchmark testing is the fastest and most reliable way to gauge the performance and throughput of the scanner(s) you are considering.

Meeting Peak Volume Demands

The equipment's rated speed and daily volume limits are primary considerations for "sizing up a scanner" for your application. It is even more important to base equipment selection on the scanner's ability to meet peak scanning loads, if workflow is concentrated in selected days or hours of the week. And since labor has a dramatic impact on ownership costs, many organizations need to consider whether to handle their daily paper volume with fewer operators and higher-volume scanners, or with more operators and lower-volume scanners. The answer depends on the organization's approach to labor, operational and capital costs.

History shows that many organizations underestimate document scanning volume requirements prior to purchase. Consequently, the selected scanner is stretched to its operating limits from the outset, unable to accommodate scanning load increases. That's why it's important to get an accurate picture of total and peak volume demands before a purchase decision is finalized.

Another way to evaluate long-term performance and cost of production scanners is to compare each unit's published Mean Time Between Failure (MTBF) rating. This rating makes it possible to assess a scanner's ability to meet operational expectations over time, given anticipated daily duty cycles.

Contributions to Costs – Segmenting the Document Capture Process

Regardless of the physical nature of the documents that need to be scanned, the strategic importance of document capture resides in the information that needs to be processed, stored, accessed, managed and distributed throughout the document lifecycle.

Document scanner functionality provides the means to turn raw data into a strategic corporate asset. In doing so, the right scanner can make it possible to lower costs, reduce cycle time and enhance the value of the captured images in subsequent uses of the information.

The process can be segmented into three stages:

- Document preparation
- Scanning and image enhancement
- Post-processing including quality assurance (QA), indexing, output and error correction

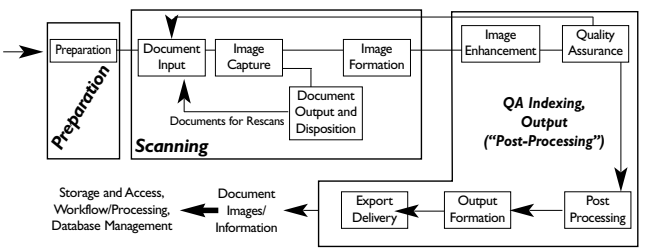
Each stage contributes to the equipment's total ownership costs, which consist of the acquisition and maintenance costs of the equipment, labor, and the costs associated with space allocations to perform each function.

Conservative estimates hold that over a three-year period, labor represents as much as 70 percent of the total ownership costs to prepare, scan, index and output documents, and perform QA. Approximately 80 percent of this labor is evenly divided between document preparation and post-processing activities. From this, it's apparent that document scanners – which reduce document preparation and eliminate rescans – have the ability to generate substantial reductions in ownership costs over the life of the equipment.

At the same time, it's important to note that capital and maintenance costs contribute as little as ten percent to the true cost of scanning. This strongly suggests that the purchase price of equipment is much less significant than the scanner's ability to contribute to process and productivity improvements over time.

Document Capture Process

A Simplified View of "Generic" Document Capture Process



Process Goals (Value Adding Opportunities)

- Lower costs
- Improve speed/reduce cycle time
- Improve value in subsequent image uses (workflow, storage, viewing, distribution)
  - Image quality, file sizes, color

Measuring Scanner Productivity

When assessing productivity, the size, weight and uniformity of the documents are among the factors that need to be considered. The unit's paper handling capabilities, image quality and scanner reliability also have a considerable impact on scanner productivity.

Document preparation can involve as much as 40 percent of the total labor in scanning, especially in document capture environments, which are subject to a great deal of time-consuming sorting, batching and other manual intervention – such as removing staples. Documents that are crumpled, torn, or folded present special challenges – as do "odd"-size documents such as memos, "sticky" notes, folders and envelopes.

Under these conditions, the scanner's document handling system must have the ability to process these exceptions without jamming.

Document Feeders. In recent years advancements in document feeder technology have resulted in superior paper separation capabilities that permit batching of mixed document weights and sizes in stacks of up to hundreds of pages. In evaluating automatic document feeders, prospective buyers should compare the occurrences of document multi-feeds and skewing among the same or similar document batches. In certain auto-feeder designs buyers need to be wary of inefficient "urging" or feed rollers that exhibit premature wear after feeding a low number of documents. When the urging or feed roller fails to grab papers in the stack, users need to continually intervene in order to feed the paper.

Multi-feed detection based upon document length has proven to be the most reliable technique for applications that scan documents of the same length. Partially overlapped documents are detected because the combined length exceeds acceptable pre-set limits.

Multi-feed detection can also be based upon document thickness, using mechanical or ultrasonic technology. Mechanical multi-feed detection is well suited for applications where document thickness (weight) is constant. However, mechanical components come in contact with the scanned document, which can introduce additional document skewing and ripping, as well as increased paper jams.

Most industry watchers believe that ultrasonic multi-feed detection technology still requires refinement in its implementation. It is often positioned as a solution for applications that need to accommodate varying document lengths and thicknesses. Some manufacturers' implementations of this technology still produce false positive conditions that can seriously affect productivity.

Ultimately, multi-feed detection cannot make up for a poor feeder design. Instead, multi-feed detection can complement a scanner with excellent paper handling.

Paper Transport. In addition to feeder design, document transport systems contribute to overall productivity by accommodating a wide range of paper weights. Enhanced transport mechanisms play a key role in minimizing pre-sorting, misfeeds and interruptions.

Buyers should carefully scrutinize the length and shape of the transport system's paper path. Field studies confirm that a short "J" or "C" path, where the leading and trailing edge of the document remain visible to the operator, offers superior scanning reliability over a wide range of document sizes, thicknesses and surface quality.

At the same time, paper transport components should be characterized by a high co-efficient of friction, as well as excellent wear and contamination resistance. Rollers should provide sufficient force to drive even the thickest of documents without introducing unequal pressures across the paper path. The path should provide an ultra smooth, scratch-free surface that is highly durable and resists static build-up. As measured by surface resistivity, static resistance as low as 10^6 ohms can be achieved on selected scanner models.

Image Enhancement. Today, production scanners can significantly vary in their ability to match or improve the clarity of the original document. At a minimum, buyers should demand on-board dynamic thresholding capabilities that take advantage of all the scanned data, to ensure optimal image quality and reduce the need for expensive post-processing boards.

Dynamic thresholding provides uniform image quality from batches of intermixed documents, eliminating time-consuming presorting by background color, paper texture, print density or quality. This important functionality makes it possible to process poor-quality and low-contrast images, such as carbonless multi-part forms, faxes and multiple-generation copies. Technical specifications should also include the presence of "intelligent filters," which lower image noise and minimize artifacts on white-to-black and black-to-white transitions.

Often, improved compression algorithms result in smaller file sizes and more legible images by storing only critical visual information, while dropping out extraneous input such as background color.

Scanner Optics Technology and Image Resolution. Scanner optics also play an important role in determining the level of image quality which impacts the need for rescans. The scanner's camera should provide extremely sharp edges and reduced noise for optimal readability. Most bitonal scanners used in document imaging applications operate at 200, 240, 300 or 400 dots per inch (dpi), with lower resolutions providing faster scanning speeds. For many applications, 200 dpi provides acceptable resolution in order to optimize throughput.

However, an important distinction must be made between capture resolution and output resolution. For example, image processing algorithms (scaling) are used selectively to output images at 300 dpi from images originally optically captured at 200 dpi, to enhance resolution specifications. In many of these cases, the higher output resolution (300+ dpi) will produce an inferior image compared with an image captured and output with native optical resolutions (200 dpi). Often, the quality of a system's optics and sensors will provide sharper, clearer output captured at lower resolutions than a scanner that relies on software to boost its resolution values. That's why comparing scanners based upon dpi resolution alone can be misleading.

Color scanning creates 24 times more visual information then bitonal scanning. Elements such as graphics, stamps, logos, signatures, and highlights are often more readable in color than in bitonal, even at a lower dpi. Color scanning resolutions of 100 and 150 dpi are considered highly comparable to bitonal resolutions of 200 or 300 dpi in both quality and image size. This is possible at a lower resolution due to the bit depth. sixteen million colors are displayed in the color image versus only two colors in the black and white image.

In terms of camera system reliability, contact array designs reside very close to the paper, increasing the incidence of paper jams and the potential for reduced image quality. It's generally accepted that lens reduction systems offer enhanced throughput and consistent image quality.

Ergonomics. Special attention should be given to scanner ergonomics. Controls, feeder and the exit hopper should be within easy reach to help operators work efficiently and comfortably, based upon their work environments. For example, top-feed units are more appropriate for walk-up applications, whereas dedicated scanners used by a single individual over the course of a workday should include document feeders positioned at the seating level.

In dedicated scanning applications, operators should be able to reach both the input and output hopper without standing. In either case, ergonomics deserve special consideration when there is a need to accommodate individuals with disabilities who may routinely use the scanner.

Also keep in mind that in some scanner designs, documents are not in the order they were scanned in the output hopper. In these cases, a separate inverter must be purchased which flips the documents to preserve their order.

Manual and Automatic Indexing. Since scanned documents will often need to be electronically retrieved as part of routine workflow and database management, robust indexing and post processing capabilities become a priority consideration in selecting a production scanner. With manual indexing, an operator views the scanned image and enters data from the document image into a data capture application.

As document capture volume increases, automatic indexing offers the added capability to minimize data entry labor costs and helps speed document processing, which can help to offset its higher capital investment costs.

Ideally, image size, bar code data and a sequence number should be made available to the scanner application. Patch reader options are becoming increasingly popular for their ability to automate the document batch structure according to patches printed on documents or header cards. Bar code and patch code recognition can be implemented on either the host or the scanner, depending on available scanner functionality and recognition rate comparisons.

Scanner Integration. Since the scanner is a single element of document capture, scanning software and related interfaces must have the capability to sustain required throughput rates. The Small Computer System Interface (SCSI) has become the de facto standard for scanner system integration. SCSI works with standard ISIS and TWAIN drivers to facilitate integration with host application software. SCSI adapters are relatively inexpensive. Video interfaces are generally found in older scanner models. A video interface requires a relatively expensive host-compatible card to process the images from the scanner.

With the scanner's SCSI interface, compressed images are buffered to overcome bandwidth constraints. It's important that sufficient memory is configured in the scanner to prevent potential throughput bottlenecks. Otherwise, there could be a noticeable drop in speed during a large, batch scanning job.

Compute Your Real Cost of Scanning

By inputting the following data you can automatically calculate ownership costs based on scanner jam rates, and reliability estimates for the scanner model you are currently using or are considering purchasing. It will be helpful if you have already collected the following information:

- Number of standard-size sheets scanned each day
- Scanning rate per minute in portrait orientation
- Purchase price of the scanner
- Purchase price of the computer and interface cards needed to operate the scanner
- Cost per hour of the scanner operator
- Monthly charge for facilities
- Annual cost of repairs

PREPARATION

Capital:	Tables, paper joggers ($ _____)
Maintenance:	None
Labor:	____ people @ $ _____ per year
Space:	_____sq.ft. @ $ _____
Three-year total:	$ _____

SCANNING

Capital:	Scanner, PC ($ _____)
Maintenance:	$ _____per year
Labor:	____ person @ $ _____
Space:	____ sq.ft. @ $ _____
Three-year total:	$ _____

POST-SCAN PROCESSING

Capital:	____ PCs, Printer, Software ($ _____)
Maintenance:	$ _____ per year
Labor:	____ people @ $ _____
Space:	____ sq.ft. @ $ _____
Three-year total:	$ _____

Making the Right Choice

Ultimately, document input requirements will determine whether your enterprise is best served by installing fewer, centralized ***High-Volume Production*** scanners or many decentralized ***Mid-Volume Production*** or ***Low-Volume Production*** units. In either case, scanner reliability and service support will play a major role in meeting productivity and quality expectations.

Since it is rarely practical to QA more than a small sample of the documents being scanned, users must rely on the consistency of the scanner's subsystems to provide predictable results, regardless of the inevitable variations in paper size, thickness and the quality of the originals. Even the most automated scanning systems involve extensive document preparation and monitoring to ensure that critical documents are available for subsequent processing, viewing, distribution, and storage.

Vendors and scanners should be evaluated on their ability to improve the document capture process as part of an effective document workflow and warehousing solution to preserve the past, protect the present, and enable the future.

For more information or assistance regarding your scanning needs, please call Kodak at 1-800-243-8811, visit our Web site at www.kodak.com/go/docimaging, or contact your Authorized Reseller of KODAK Document Imaging Products.