DOCUMENT RESUME

ED 353 988 IR 054 364

AUTHOR	Willis, Don
TITLE	A Hybrid Systems Approach to Preservation of Printed Materials.
SPONS AGENCY	Commission on Preservation and Access, Washington, DC.
PUB DATE	Nov 92
NOTE	71p.
AVAILABLE FROM	Commission on Preservation and Access, 1400 16th Street, N.W., Suite 740, Washington, DC 20036-2117 ($10 prepaid).
PUB TYPE	Viewpoints (Opinion/Position Papers, Essays, etc.) (120) -- Reports - Evaluative/Feasibility (142)
EDRS PRICE	MF01/PC03 Plus Postage.
DESCRIPTORS	Access to Information; Comparative Analysis; *Costs; *Evaluation Criteria; Information Storage; *Microreproduction; Optical Scanners; *Preservation; Standards; Technological Advancement
IDENTIFIERS	*Digital Imagery; *Printed Materials

ABSTRACT

Intended to stimulate thought and discussion, this report compares micrographics and digital imaging as tools for the preservation of printed materials. The topics covered include: (1) the advantages and disadvantages of each technology; (2) trade-offs involved in selecting one technology over another; (3) benefits of using a hybrid approach; (4) whether the page should be captured first to film and converted to digital, captured digitally and converted to film, or whether the two can be done simultaneously; (5) the options for converting from film to digital and back again; (6) cost factors, including how to maximize image quality while minimizing cost; (7) the roles of ASCII text and OCR (optical character recognition); (8) resolution issues for each technology; and (9) standards. It is concluded that microfilm will preserve printed materials very well and that the equipment needed to transfer this material to other media will be available for centuries; and that optical storage can be considered on a selective basis provided there is a plan to recopy the media prior to any substantial degradation and before the technology becomes obsolete. It is recommended that, for the longer term, practitioners should immediately begin planning for, and designing, the hybrid archival preservation system of the future. It is suggested that such a system could combine the strengths of micrographics with digital imaging, which contributes access, distribution, and transmission strengths. A discussion of digital imaging resolution, a summary of alternative storage possibilities, data storage costs in a variety of formats, a comparison of film and digital costs, and a list of resources for equipment performance standards are appended. Examples of images copied using different media are also provided. (KRN)

The Commission on Preservation and Access

A Hybrid Systems Approach to Preservation of Printed Materials

by
Don Willis

Prepared for the Commission on Preservation and Access
November 1992

1400 16th Street, N.W., Suite 740, Washington, D.C. 20036-2217 • (202) 939-3400

A private, nonprofit organization acting on behalf of the nation's libraries, archives, and universities to develop and encourage collaborative strategies for preserving and providing access to the accumulated human record.

This report, prepared at the request of the Technology Assessment Advisory Committee (TAAC), is one of a series alerting the Commission on Preservation and Access and others to developments and possibilities within the context of national and international initiatives for preservation of and access to information printed on disintegrated paper and other substrates.

The paper was subjected to a pre-publication review by TAAC members, although it remains principally the work of Don Willis, Vice President, Electronic Product Development, University Microfilm International. TAAC hopes that this publication will spark additional thinking, discussion and progress regarding reformatting technologies and that it will contribute to our collective understanding of how preservation and access needs can be addressed by emerging technologies.

NOTE: Definitions of most technical terms appear in M. Stuart Lynn's glossary, *Preservation and Access Technology/The Relationship Between Digital and Other Media Conversion Processes*, published by the Commission on Preservation and Access, August 1990.

Acknowledgments: *The author would like to thank Heinz Dettling, Bob Mottice, Rick Closner, John Brooks, and Elaine Cavin for their help and critical review, also Hosni Adra for preparing the image enhancement exhibits. The author is particularly grateful to Kaitlin Hanger, Tina Creguer and Keri Bryan for their graphic and editorial work. And finally, the author would like to thank various members of the committee for their input and direction--in particular Rowland C.W. Brown, who helped tremendously with the structure and presentation of this work.*

Published by
The Commission on Preservation and Access
1400 16th Street, NW, Suite 740
Washington, DC 20036-2117

November 1992

Reports issued by the Commission on Preservation and Access are intended to stimulate thought and discussion. They do not necessarily reflect the views of Commission members.

Additional copies are available from the above address for $10.00. Orders must be prepaid, with checks made payable to "The Commission on Preservation and Access," with payment in U.S. funds.

This paper has been submitted to the ERIC Clearinghouse on Information Resources.

The paper in this publication meets the minimum requirements of the American National Standard for Information Sciences-Permanence of Paper for Printed Library Materials ANSI Z39.48-1984.

TABLE OF CONTENTS

Introduction page 1

Comparing Micrographics and Digital Technology 1
Areas of Analysis 1
Executive Summary 2
Recommendation 5

The Issues 5

What Are the Advantages and Disadvantages of Each Technology? 5
-- Micrographics 5
-- Digital imaging 6
-- Summary 6

The Optical Disc 7
-- The improving optical disc access solution 7
-- High-capacity "permanent" storage 7
-- Optical discs: how long will they last? 8
-- Redefining "archival" 9
-- Impact of obsolescence on the digital approach 9
-- Preservation through rewrite 9

Resolution, the Key Design Element 10
-- Micrographics 10
-- Digital imaging 11

The Trade-offs in Selecting One Technology Over the Other 12
-- A film-only system 12
-- A digital-image-only system 12

The Benefits of a Hybrid-System Approach 13
-- Playing to their strengths 13
-- The hybrid end-user access system 14

Film First, then Convert ... or Vice Versa? 15
-- Filming first 15
-- Scanning first 16
-- Timing and volume, two key factors 16
-- Simultaneous scanning and filming 16
-- Digital computer output microfilm (COM) camera 17
-- Digital technology still under development 17

Options for Converting from One Format to the Other 18

ASCII Text and OCR 19
-- Extracting character code data from a page image is always an option 19
-- ASCII text--limited preservation usefulness 19
-- Other data formats 20

Image Access, Distribution, and Transmission 20
-- Access 20
-- Distribution 21
-- Transmission 21

Costs 21

Handling Pages--Little Difference in the Cost 21

Cost of a Digital Image Preservation System 22
-- Digital system implementation costs 22
-- Digital system operating costs 23
-- Contract preservation imaging costs 23
-- Optical disc drives and media costs 24

Cost of a Micrographics System 24
-- Expected workload 24
-- Micrographics system implementation costs 24
-- Micrographics system operating costs 25
-- Contract microfilming costs 25
-- Film storage and duplication costs 26

Preservation Cost Summary 26

Conclusion 26

Recommendations 27

Get Involved 27

Understand the Technology 27

Minimize Risk 27

Prepare for the Future 28
-- If the technology were available 28
-- The future digital solution 28
-- The long-range system 28
-- An optimal mix 29

Appendix A: Resolution -- A Key Design Parameter 33

The Single Most Important Factor 33
-- Designing around the amount of data 33
-- How high resolution affects cost 33

How Much Image Resolution Is Required? 33
-- Capturing small type sizes 33
-- Halftone resolution 34
-- Archival resolution 34
-- Greyscale scanning can improve page quality 35
-- Image enhancement 35
-- Standard compression algorithms 35
-- Equating scanner resolution and film resolution 36
-- The Nyquist sampling theorem 36

Digital Image Scanners: How They Work 36
-- Binary scanners 36
-- Binary page storage requirements 37
-- Greyscale scanners 37
-- Greyscale page storage requirement 37

Printing 38
-- The laser printer 38
-- Creating a halftone 38
-- Halftone printing with a laser printer 38

Appendix B: A Summary of Storage Possibilities 39

Appendix C: Table -- Data Storage Costs 40

Appendix D: Table -- Preservation Cost Summary 41

Appendix E: Standards 42

NOTES 43

FIGURES 45

INTRODUCTION

Comparing Micrographics and Digital Technology: This paper will focus on questions about the use of micrographics and digital imaging technologies for preservation of printed materials. It will not address any of the issues involved in the preservation of sound, motion pictures, video, art, or color images. The author is aware that other document preservation issues exist; however, it was felt that these two technologies were of most interest to the preservation community at this time. Topics to be covered include:

- What are the advantages and disadvantages of each technology?
- What are the trade-offs involved in selecting one technology over the other?
- What are the benefits of a hybrid approach?
- In a hybrid system, should the page be captured first to film and converted to digital, or vice versa; or can it be done simultaneously?
- What options are available for converting from film to digital and back?
- What are the cost factors; how does one maximize image quality while minimizing cost?
- What role should ASCII[1] text and OCR (optical character recognition) play?
- How can the required resolution be determined, and what are the resolution issues with each technology?
- What standards should concern the practitioner?

Areas of Analysis: There are three primary areas of analysis in comparing digital electronic image systems to film-based systems for preservation: document capture, storage, and access. In capture, the analyst will be concerned with the capture mechanism, resolution, quality of the captured image, acquisition speed, system cost, operating cost, and indexing requirements. In storage, the concerns are media permanence, media refresh requirements, technology obsolescence, drive cost, media cost, interchangeability of media, reliability, performance and access tradeoffs. Finally, with regard to access, the designer must examine retrieval capability (both searching and browsing), retrieval speed, transmission and distribution capability, and retrieval quality. Micrographics and imaging technologies can complement each other and best address these concerns together in the well-designed preservation system.

---

[1] American Standard Code for Information Interchange

This paper will survey micrographic and digital technologies in light of the issues and concerns defined above. The objective is to arrive at short and long-term recommendations for developing document preservation systems based on these technologies.

Executive Summary: Based on a review of the technology, our findings are:

o Design objectives are extremely important: The preservation systems designer must identify the objectives of the preservation system in detail. For example, if practitioners desire to preserve a faithful reproduction of the document, do they want the page as it currently exists complete with its discoloration due to age and water stains, or do they desire a cleaned up page, similar to what was originally published? Obviously, an image can only be cleaned up by using electronic technology, so system requirements have a definite impact on the technology that must be used.

Other important system design criteria include the volume of the workload, quality required, methods for storing and accessing the documents, frequency of access, urgency of access, response-time requirements, condition of the documents, and page sizes[2].

o A micrographics-based preservation system is a generally acceptable solution here and now for most printed materials. It is a mature technology with widespread familiarity and a large installed base. High-quality film created and stored according to standards will last up to 500 years.

o Centralized master vaults already exist where over 3 million rolls of film masters are stored in secure, climate-controlled conditions for only about $1.00 per reel per year.

o Microfilm's major weakness is its inadequate access and distribution characteristics.

o Although microforms are currently a relatively inexpensive preservation medium for printed materials, costs for this type of solution will increase at five to ten percent per year due to the increasing cost of labor.

o Micrographics cannot be considered acceptable solution for all preservation needs; for example, it is not ideal for preserving l. h-quality greyscale images, color images (e.g., artworks), sound recordings or full motion video. In these areas, digital technologies are the only reasonable alternative.

o It can be twenty times more expensive to store 9 X 5 inch archival resolution page images on optical disc than on 35mm film.

---

[2] Throughout this document (unless otherwise noted) the page size used is a conservative measurement for the typical journal page of 8.5 X 11 inches or 93.5 square inches. Since the typical book is only 5 X 9 square inches or 45 square inches, the storage space needed for a digital representation of book pages at any resolution is about half of that required for the journal size page.

o For digital preservation systems, productivity increase will be brought on by technology advances, and these advances are expected to accelerate rapidly over the next several years.

o There are no forms of digital storage currently on the market that would be considered archival according to the traditional definition.

o Write-once optical disc could be considered permanent[3] but not archival. The reason is not the longevity of the media -- it's the fact that the technology becomes obsolete. Even if the media were to last 50 years, chances are there wouldn't be a drive available to play it.

o Perhaps when referring to digital storage media, "archival" needs to be redefined as the ability to recreate an exact copy from the original medium before it degrades or the technology necessary to read it becomes obsolete.

o Assuming that refreshing of media (recopying) would be cost justified by the increase in capacity and/or reduction of cost of the new media, a key question preservationists must answer is, "Is a solution acceptable which requires the media to be recopied onto more advanced media every "N" years in order to keep up with advancing technologies?" If so, who would be in charge of assuring that the conversion was carried out on schedule? This whole topic could be the subject of a new paper.

o A digital image based preservation system is the most promising future solution for printed materials. It is a rapidly changing technology in quality, speed, and economics. Its major weaknesses are that the technology is fairly new, has high data-storage requirements, and lacks *proven* archival storage capability.

o Digital imaging technology will increase in functionality and decrease in cost for the foreseeable future. Many experts believe that an all-digital system will provide the most economical future preservation solution. In fact, if one were to do a five year present value analysis of a micrographics based versus a digital image based preservation system today, factoring in the costs of access and distribution, the digital system would most likely prove to be the least expensive alternative.

o Access to the preserved materials is a key benefit of the digital image preservation system. Access can be through a separate database of indexes, abstracts and indexes, full-text search on the ASCII portion of compound documents, or by browsing through the database item by item.

o With digital technology it will no longer be necessary for the researcher to travel to where the preserved materials are physically located; access to historic collections throughout the country can be as close as the nearest computer or printer.

---

[3] Continuing or enduring without fundamental or marked change.

o Efficient access to the preserved collections has the potential of allowing the institution to self-fund some of the preservation costs through revenues generated from the improved access to the archival collection.

o An inexpensive solution to preservation has been explored in a pioneering project of Cornell University. They have used digital scanning at 600 dots per inch (dpi) binary to create high-quality copies on acid-free paper. The idea is to create a permanent, not archival, paper copy that can go back on the shelf -- preservation reformatting.

o A hybrid system, one that combines both film and digital imaging, could well offer the best overall design for current preservation needs. Micrographics provide a relatively inexpensive, high-quality archival storage medium. Digital imaging contributes access, distribution, and transmission strengths. It should be noted that in the near future, most national service bureaus will have the capability to transfer from one technology to the other, so the practitioner need not design the full hybrid capability into the local system.

o A hybrid system can be implemented with today's technology by filming first and scanning some or all of the film to enhance access to the preserved collection. We will designate this as the "film-first archival preservation system."

o The latest possibility for implementing a hybrid system is through filming and scanning simultaneously. New belt-fed combination duplex scanner/filmer image capture devices were introduced at the 1992 AIIM show by Bell & Howell and Kodak. These devices could be used on non-brittle documents. As far as processing goes, this type of system suffers from some of the same limitations as the film-first system which will be discussed later.

o The "scan-first archival preservation system" is rapidly becoming an acceptable alternative for the preservation system designer. By scanning first, each page can be decomposed into separate areas of text, line art, and halftones. Each of these will be electronically processed independently to maximize overall page quality. By scanning in greyscale and enhancing the digital data prior to creating film, it will be possible to create higher quality film than can currently be created using light/lens methodology.

o Scanning first will also allow more intelligent retrieval aids in bar code format or blip marks to be recorded onto the film so that retrieval can be automated.

o Digital imaging allows end-users to obtain higher quality printed copies than micrographics. Each copy will be a first-generation copy. As with music on a compact disc, there is no degradation during usage. Because of the aforementioned, the scan-first archival preservation system will be more cost-effective to build and operate than any other type of preservation system once all the technology is available.

o Resolution is the key design parameter for a digital image preservation system (see **Appendix A**). We've defined various levels of resolution referred to in this paper as follows:

-- "Archival resolution" is defined as the resolution necessary to capture a faithful replica of the original document, regardless of cost.

-- "Optimal archival resolution" is the lowest resolution that will completely satisfy the archival image objectives defined for the system.

-- "Adequate access resolution," on the order of 300 dpi binary, is defined as the resolution sufficient to capture about 99.9 percent of the information content of the page.

o Microfilm is "resolution-indifferent". Each frame of film can store high-quality images with equivalent digital resolution of about 800 to 1,000 dpi with about 8 - 12 levels of greyscale.

o Digital imaging is "resolution dependent": the higher the resolution requirements, the higher the cost and complexity of the system.

o The above suggests a second question pertaining to resolution that must be answered if we are to accurately evaluate our alternatives. It is "should film standards, which primarily measure the high contrast components of a reproduction, be used to measure digital reproducibility?" Do we want to have perfect print or a high-quality copy of the entire original including halftones.

Recommendation: Currently, practitioners choosing microfilm for a preservation solution can feel confident that their printed materials will be adequately preserved and that even in the next century or beyond the technology will be available to transfer this material to other media if desired. This is true because of its accepted archival nature, and the fact that one only needs a lens and light to read it. Optical storage can be considered for preservation on a selective basis provided there is a plan to recopy the media prior to any substantial degradation. For the longer term, practitioners should immediately begin planning for, and designing, the hybrid archival preservation system of the future. The continuous and accelerating improvements in electronic imaging and optical disc technology will be the key to solving preservation problems.

THE ISSUES

What Are the Advantages and Disadvantages of Each Technology?

Micrographics

Advantages: As a storage medium, microfilm is durable and relatively inexpensive. Standards for creating, processing, storing, and reading microfilm are well known; the equipment necessary to read microfilm is not likely to become obsolete (all that is needed is light and magnification); microfilm copies are recognized as legally acceptable substitutes for original documents; microfilm can theoretically store high-quality greyscale images inexpensively; and it is a recognized archival medium (ANSI IT9.5-1988, ANSI PH1.67-1985) with a large installed equipment base. See **Figure 1**.

Disadvantages: Film can become scratched when handled; consequently, archival film is usually stored in a vault, and only copies are distributed for general use. Each generation or succeeding copy loses resolution (about ten percent). In addition, most micrographics reader/printers must access the film manually; reader/printer blowbacks (printouts) are of poor quality; film creation variables are difficult to control; film quality can only be determined after filming is complete; and bad pages must be re-filmed and spliced in.

In addition, there is no way to selectively tune the input process to maximize quality based on page content. Some preservation projects require filming two exposures of certain pages--a high-contrast exposure to effectively capture the text and a low-contrast exposure to capture photographs more faithfully. Even with this approach, certain color combinations don't photograph well, such as black print on a red or blue background. (Some preservation microfilmers have developed a special film-processing chemistry that improves the tonal range of greyscale images while preserving the contrast--in essence giving the user the best of both worlds -- greyscale and text). Finally, the practitioner must be aware that most of the microfilm produced by the typical service bureau for records management does not meet preservation standards.

Digital imaging

Advantages: The digital image format offers ease of access; excellent transmission and distribution capabilities; electronic restoration and enhancement; high-quality user copies; and automated retrieval aids. Notice that the primary focus is on improving user quality and providing better access to the information. See **Figure 2.**

Disadvantages: The technology is relatively new; a digital image, displayed or printed, is not yet acceptable as a legal substitute for the original; standards are lacking in many areas; digital storage is not considered archival -- it requires continuous monitoring and eventual or periodic rewrite; the drive systems will inevitably become obsolete; there are relatively high but rapidly declining storage costs; the cost to store high-resolution archival images increases as the quality increases; and greyscale images require even more storage space.

Summary

Micrographics: A mature technology, generally accepted for preservation of printed materials. High quality and low cost. Major weakness -- inadequate access and distribution characteristics.

Digital Imaging: Most promising future technology for preservation of printed materials. Rapidly evolving in quality, speed of access and economics. Major weaknesses -- the technology is fairly new, data storage requirements for archival quality images are high, it lacks standards and is not a proven archival storage media.

The Optical Disc

The improving optical disc access solution: Access is the other side of the preservation coin. It is one thing to preserve a corpus of knowledge for future generations; it is another, and completely different objective, to provide researchers access to preserved materials in a way that will not damage them. In reflecting on this dichotomy, Bill Nugent, a visionary in the field of imaging and optical disc technology, says, "...[T]he dual objectives of the preservation of materials and providing ...public access to them are opposed to each other. Preservation generally means a strictly controlled physical environment, watchful custodial care, and limited public usage. High public usage generally means accelerated wear and deterioration. But page images preserved on digital optical disc or in a hybrid system can now meet both objectives without conflict, since no wear results from the low-power laser beam used to read the data from the disks." [1] Clearly, optical disc, used in a hybrid system in a hierarchical fashion, fulfills its access role quite effectively.

In addition, the fact that researchers will no longer have to travel to the physical location of the collection, the increased ability to gain access to multiple collections simultaneously, the ability to accurately and quickly retrieve very selective information, and, finally, the ability to have access to high-quality copies of historic documents are just not possible with any media but electronic. Since this increased access capability adds value to the research process, it has the potential to allow the institution to self-fund some of the preservation costs through revenues generated from charging for this improving access to these archival collections.

High capacity "permanent" storage: The optical disc was one of the primary technologies that made digital imaging practical. Digital images require huge amounts of storage space. The optical disc promised high-capacity, permanence, removability, and random access -- all at an inexpensive price. The advantages of the optical disc as a storage technology are listed in **Figure 3.** Since the optical disc is read by a laser beam, and since its metallic surface is encapsulated in plastic or glass, it has high resistance to wear during use.

All numbers contained within [ ] refer to endnotes.

There are several kinds and sizes of optical discs. The one usually discussed for preservation is the write-once-read-many (WORM) disc. It is written with a laser beam that burns holes into its metallic surface. Once data is written to the disc it cannot be erased. If an error is made and the data must be rewritten on the disc, it is rewritten in a new area, thus leaving an audit trail.[4]

Other types of optical discs include read-only memory (e.g., CD-ROM and the videodisc) and the newest member of the family: Erasable. The erasable optical disc is viewed primarily as a replacement for magnetic tape and magnetic disk. Since it can be erased and rewritten, it is not usually considered for archival storage purposes.

The CD-ROM and videodisc are primarily distribution media; however, they have the same characteristics for longevity, removability, and error correction as their write-once cousins and could be used in an overall hierarchy of storage for effective storage of preservation documents. This is particularly true with the introduction of the write-once CD-ROM, which because of the low cost of the media and the fact that it can play in a standard CD-ROM drive, should be very attractive for use as a preservation access media.

Optical discs: how long will they last? Bill Nugent defines optical disc longevity as follows:

"Longevity is the expected duration between the time of manufacture of an optical disc and the time one of its important parameters degrades to a point where the disc becomes unsuitable for use or to a measurable point pre-defined as "end-of-life" for that parameter. An example would be a disc's bit error rate (BER)[5] degrading to 1.0 X 10E-04, a defined end-of-life point for 5.25 inch write-once optical disks." [2]

He says that by conducting a series of accelerated aging tests, one can statistically determine an expected end-of-life for an optical disc based on the increase in the bit error rate. Once determined, the bit error rate of each disc can be monitored to predict approaching end-of-life and allow the disc to be copied while its integrity is still guaranteed. Since optical discs contain two levels of error correction, discs in the early stages of degradation can be recopied with no loss of data.

Longevity is critical in preservation applications. Optical disks will not be comfortably accepted (for archival storage) until longevity, decay rates, the physical nature of failure mechanisms, and a strategy for rewrite based on scheduled monitoring using prescribed test procedures (or scheduled rewrite procedures) have been established.[3]

[4] A manual or computerized record that can be used to trace the type and origin of transactions affecting the contents of a document, record or file.

[5] Measurement of the number of bits of data found to be in error when information is read off a storage medium.

Redefining "archival": When one thinks of defining archival, the definition "preservation of a document for about 500 years" comes to mind. This definition works well for information that can be interpreted by the eye, because the eye has remained the same for hundreds of thousands of years. However, technology advances rapidly. The information stored in electronic format must be interpreted through computers or computer peripherals for it to be intelligible by humans; however, two factors influence the ability to gain access to this information: the permanence of the media and the life of the technology needed to provide access to the information. The fact that digital storage media may last for 100 years or more has little meaning in and of itself. In this case, *"archival" should be redefined as the ability to recreate an exact copy from the original medium before it degrades or the technology to read it becomes obsolete.*

Impact of obsolescence on the digital approach: The National Archives, in its report "Preservation of Historical Records," claims that optical discs can never be used for permanent (I believe they mean archival) storage. The Archives is concerned about the problem of obsolescence. They cite as an example the 1960 census, which was the first to be automated. In 1970 archivists discovered there were only two computers in the world that could read the 1960 census data. One was in the Smithsonian, the other in Japan. We supposedly know less about this first "automated" census than we do about the census of 1860, 100 years prior.[4]

Obsolescence is a key concern for the designer of any digital image system. The fact that the storage device will become obsolete will require that the media be recopied every five to ten years.

Preservation through rewrite: The practitioner can monitor the media as suggested by Nugent, or adopt a policy of scheduled rewrite. There are those who feel that whichever strategy is employed, rewriting the prior generation of digital storage media onto the next generation will be cost effective because of advances in technology. However, by using the concept of the hybrid system and employing film as the system archive, the need for this rewrite (refresh) cost could be reduced or completely eliminated from the lifecycle of the system. After all, film, as a storage media, is still less expensive than optical disc, and even though the archival film needs to be stored in a vault, these storage costs will remain less than the digital media refresh costs for some time to come.

Assuming the concept of the storage hierarchy is applied within the context of the hybrid system, only a small percentage of the preserved documents (the most frequently and most recently used) will be in digital format at any given point in time. This could substantially reduce the preservation system operating costs.

A final very real concern with the need to effect preservation through rewrite is that in tough economic times refresh costs could be cut from the budget, or for whatever reason, a policy of selective rewrite, or censoring, could be adopted. Can we really rely on those who will follow us to assume the recopying responsibility?

Resolution, the Key Design Element

Micrographics

Film resolution: Film resolution is typically defined as the ability to render visible fine detail of an object; a measure of sharpness, it is expressed as the number of line-pairs per millimeter (lppm)[6] that can be "resolved". A line-pair is one black and one white line juxtaposed. A series of line-pairs is said to be resolved if all lines in an array of line-pairs on a test target can be reliably identified. Film resolution is measured by photographing several test targets, and under a microscope, determining the smallest pattern on which the individual lines can be clearly distinguished.[5] See **Figure 4**. Research Libraries Group specifications require that a resolution target be part of the initial sequence of frames for each book on a film reel, and that the measured resolution be about 120 lppm, or a ten target.[6]

Effective film resolution: Theoretically, microfilm is capable of storing resolutions of 1,000 lppm, but this theoretical limit is actually never achieved because even the best microfilm cameras operating under ideal conditions are limited to about 200 lppm. And, due to variations in lighting, exposure control, lens quality, focus, development chemistry, camera adjustment, vibration, and other variables in a production environment, high-quality 35mm 12X film is usually imaged at an effective resolution of about 120-150 lppm (The RLG standard identifies any resolution above 120 lppm, at a 12X reduction, as being excellent). This effective film resolution equates to a digital binary scanning resolution of approximately 700-900 dpi. It will be a few years before cost-effective digital image systems capable of handling this level of resolution are available on a production basis. (See **Appendix A**)

Film is resolution-indifferent: A single frame of film can store an image at the maximum possible resolution for the film/camera combination being used. Film does not exact a premium for maximizing resolution. On the other hand, the cost of storing high-resolution digital images on any medium except film increases linearly as the resolution increases. This occurs in the digital image because with higher resolution more data points are required to accurately preserve the fidelity of the image. More data points demand more memory for storage. Film, on the other hand, is resolution-indifferent.

Film integrity: Archivists are comfortable preserving materials on microfilm, because they know that--assuming the film is manufactured, processed, and stored according to established standards--they are creating a permanent record that will possibly last hundreds of years.

---

[6] Line-pairs per millimeter or lines per millimeter is a measurement of resolving power. The resolution test pattern is made up of black lines on a white background: the black lines and the white spaces are of equal width. A test pattern is said to be resolved if all five lines in both directions can be clearly differentiated.

Digital imaging

Background: Digital imaging technology is viewed by many as a replacement for microfilm; however, that perception is not completely accurate. It will be a few more years before optical disc will be a cost-effective storage medium replacement for film. In general, most people are familiar with micrographics. Conversely, many people are unfamiliar with the intricacies of digital imaging technology.

Digital image resolution: Digital image resolution is commonly defined as the number of electronic samples (dots or pixels) per linear unit measure in the vertical and horizontal scanning directions. The term pixel refers to (picture elements). A digital image is analogous to an electronic photograph. It consists of a series of pixels that can be reassembled in the proper sequence to reconstruct the original page. These pixels are represented in computer memory by a digital code. Most image scanners commercially available range in resolution from 200 to 600 dpi and are referred to as bitonal or binary scanners because the pixels can only be represented as either black (0) or white (1). If the scanner captures greyscale pixels, then the quality of any continuous tones or halftones on the page will be more accurately captured. Greyscale pixels reflect the value of the light being reflected off the page and, for 8 bit pixels, are represented by a number on a scale between pure black (0) to very white (256). The number (i.e., density) of dots is governed by the resolution of the digital image scanner. The higher the resolution, the higher the fidelity of this recreated representation.

Because these digital dots (pixels) are very small, a great deal of them are required to recreate the image. For example, at a resolution of 300 dpi, 90,000 dots per square inch are generated. This is why large amounts of storage space are required to store high-quality image data.

o For this paper we've defined various levels of resolution referred to as follows:

-- "Archival resolution" is defined as the resolution necessary to capture a faithful replica of the original document, regardless of cost. Currently this seems to be on the order of 600 dpi with eight bits of greyscale, it may well turn out to be higher

-- "Optimal archival resolution" is in effect the highest resolution that technology will economically support at any given point in time. It is aimed at achieving the optimal balance between minimal system cost and maximum image quality.

-- "Adequate access resolution," on the order of 300 dpi binary, is defined as the resolution sufficient to capture about 99.9 percent of the information content of the page. It is not suitable for preservation; however, it is generally acceptable for most information access requirements.

Digital imaging is not resolution-indifferent: As resolution increases so does the amount of data captured. The time required to scan and process the image, the quality, fidelity, and amount of storage space required to store the image also increase in direct proportion to increasing resolution. System resolution objectives must be examined in depth during systems design. Design trade-offs involving quality versus cost will influence every decision regarding resolution. For a detailed explanation of resolution issues, see **Appendix A.** It is important to determine exactly what the system's objective is so the system designer can determine the minimum economical resolution that completely satisfies the quality objectives. The idea is to maximize quality while minimizing cost.

The Trade-offs in Selecting One Technology Over the Other

A film-only system: The trade-offs involved in implementing an all-film preservation system at this time are: a) the film produced must be of the highest quality balancing high-contrast text with a wide range of greytones; and, b) typically, in film systems, very little attention is paid to indexing and creating automated retrieval capabilities; therefore, if the film is ever converted to digital, the access methods will have to be created at that time.

Designing a preservation system based on micrographics technology alone requires that all standards for the creation, handling, processing, and storage of the film be scrupulously followed. Also, it's important that the film created be of very high quality with a good balance of high and low-contrast content. However, indexing the film the way a typical digital collection would be indexed will most likely not be done. Of course, the individual publication or document can be identified along with the film roll or fiche on which it is contained, but it is extremely difficult to identify articles, pages, or the relationship between the two in a film-based system. Film indexing is just something not usually done because film access is usually sequential.

The choice is to live with the inefficient retrieval characteristics and low-quality blowbacks (printouts from a reader/printer) that are inherent disadvantages of film or to add digital retrieval at a later date. This can be done; however, the newly created digital page images will have to be further indexed to take full advantage of the digital image retrieval capabilities. This means a duplication of some of the document handling work done earlier when the film was first captured, but this incremental cost must be paid in order to enhance access.

A digital-image-only system: The trade-offs involved in implementing an all-digital preservation system at this time are: a) the designer might try to economize on the system by designing to a lower resolution, thus reducing implementation and operating costs at the expense of capturing a less-than-archival image; b) the operating budget may not include the cost of rewriting the optical disks; and c) all the quality and technical issues necessary to implement an archival digital image system have not yet been resolved.

A preservation system designed around only digital image technology must be configured to solve three major problems: 1) the lack of a true archival storage capability, 2) the need to scan at high resolution (around 600 dpi or higher with greyscale), to create an archival quality image, and 3) the high but declining cost of archival resolution image storage on optical disc. The fact that digital imaging is not resolution-indifferent ı 'hat the cost of image storage will be high. For example, to store archival-quality pages .n optical disc using JPEG[7] requires approximately 2.25 megabytes (MB) of storage space (see **Appendix A**, "Greyscale scanners").

With the average 12-inch optical disc costing about $300 (in quantities), and having a storage capacity of about four gigabytes (a GB is 1,000 MB); then, 3,540 greyscale 9 X 5 inch images at a resolution of 600 dpi can be stored at a cost of $0.085 per compressed page (media cost only). This same resolution image can be stored on film for less than $0.01 per page. In addition to the higher initial storage cost, the designer will have to figure in the cost of rewriting the disks every five to ten years. This rewriting cost may well be offset by the increase in storage capacity or decrease in technology cost over time.

Thomas Bourke, a well-known researcher in the area of applying micrographics and optical disc technology in libraries, in an article entitled "Research Libraries Reassess Document Preservation Technologies," notes that the Committee on Preservation of the National Archives and Records Administration made a recommendation to the Archivist that all holdings within the Archives be preserved on human-readable film, because this mature technology will not change significantly in the future.[7]

It seems that the Archives committee has concluded, as have many experts, that today an all-digital system is still a slightly risky preservation approach. But within the near future, technology will evolve; and the policy, standards, and administrative issues will be resolved, with one likely outcome being that the hybrid preservation system would become the accepted preservation approach.

The Benefits of a Hybrid-System Approach

Playing to their strengths: The requirements of a preservation system are best met with a combination of technologies. Digital imaging has two primary strengths: 1) The capability to improve access, transmission, and distribution of preserved images; and 2) The ability to electronically enhance (clean up) images. It eliminates some drawbacks that have kept micrographics from being a more widely acc.pted document storage and retrieval technology, instead of simply a space-saving technology.[8,9]

---

[7] Joint Photographic Expert Group

Micrographics, on the other hand, is currently the only truly archival preservation media. It is excellent for providing long-term storage for massive amounts of infrequently used information. See **Figure 5**.

By taking advantage of the strengths of film combined in a hierarchical system with the access capabilities provided by digital imaging, a preservation system can be designed that will satisfy all known requirements in the most economical manner.

The hybrid end-user access system: In addition to the hybrid system designed to preserve the materials, there must also be hybrid systems that will allow access to the preserved collections. These systems could be both local and remote and will most likely be connected together via local or wide area networks. They should consist of file servers and end-user workstations.

The file servers provide access to both bibliographic catalogs that can be searched to determine where to locate items of interest and image databases containing images of the preserved documents.

The workstation (either a UNIX type system, or high-end PC 386 - 486) should be a key component in the design of any digital image preservation system. The design should focus on a distributed system based on the client/server model, where the workstations do the bulk of the work. The workstation should be used as the production engine or an end-user access station. If the system is designed in this manner then advances in workstation technology represent potential for tremendous operating efficiencies obtained by simply upgrading to the next generation of workstation processor. The benefit of doing this is that the systems designer can depend on the fact that the workstation will increase in power at the rate of about 25 percent per year, and the cost will decrease at the rate of 10 to 20 percent per year. Therefore, the price performance ratio of the entire preservation system gets better every year -- automatically.

The production workstations would be connected to the preservation system via a local area network. They are used to perform the preservation functions such as batching, scanning, indexing, controlling the creation of digital film, etc.; all of the functions required to archive the documents.

On the other hand, the end-user access workstation will allow researchers to gain access to the databases of preserved documents. The system provides access to text, digital image, and multi-media databases distributed on CD-ROM, multi-media databases of images on videodisc, online networks (such as BRS, Dialog, and EPIC) as well as a document ordering capability, facsimile document delivery, and computer-assisted film retrieval. Access to one or more preservation databases--online or on CD-ROM--will allow the user to find citations to content of interest and request facsimile printouts on a local high-quality binary printer. See **Figure 6.** In this manner the end-user system can be useful regardless of the storage media or the technology used to preserve the materials. Where copies of documents will suffice, they can be delivered in fax format within hours of the request. Researchers will save considerable time and money by not having to travel to where the preserved materials are located, thus eliminating hardships for the researcher and artificial barriers to access.

Film First then Convert...or Vice Versa?

Filming first: Within the hybrid system concept, if an institution chooses to create film as the first step in the preservation process, the system designer can choose either low or high-contrast film based on the type of material being processed and optimize the chemistry accordingly. With film there is little flexibility for handling pages differently based on content unless multiple (low and high-contrast) exposures are used for each page, or unless, through some special processing and/or chemistry, the tonal range of the film can be extended. Typically, with low-contrast film some resolution and text clarity will be sacrificed. On the other hand, high-contrast film means better text rendering with fewer grey levels. Micrographics is basically a high-contrast process.

Many experts recommend filming first and then scanning the film. Their theory is that since the light shines through the film being scanned, most of the light can be captured by the CCD (charge-coupled device)[8] scanning array, and a better image created. In hardcopy scanning, the light reflects off the page in various directions and only some portion of it is captured by the CCD array. Although more light might be captured while scanning film, this advantage is offset by the fact that the film is already a generation away from the hard copy original and has lost some of its original resolution and greylevels. Therefore, image quality is probably about the same, regardless of whether the image is scanned from hardcopy or film.

Glen Magnell, director of marketing for the Document Imaging Systems Division of the Minolta Corporation, claims that microfilm is the most efficient input medium for recording onto optical disc. Magnell says that "...scanning from microfilm is much more efficient and virtually as reliable as hardcopy scanning [emphasis added]." [10]

I would disagree. Filming first works well if the documents require little processing in conjunction with the capture process. That's because film is a linear medium, so it can only be used by one person or process at one time. When filming, the hardcopy must be processed in the exact order as it should appear on the film, and QC is only performed after the film is developed. The filming process requires a good deal of batching, rework, and splicing which makes it quite inefficient.

[8] Type of electronic component that senses light. It builds up an electrical charge in direct proportion to the amount of light registered. The electrical charge can be read out for each individual element within the array to recreate an image line by line.

On the other hand, when hardcopy is converted to digital form it is extremely easy to process the page (e.g., indexing, real-time QC (quality control), OCR, sorting, batching and parallel processing); all are inherent in the technology.

A second concern is the limited number of microfilm scanners available and their limited resolution options. Because the demand for preservation scanning from film is small, it may be necessary for the system to have a microfilm camera custom-modified to meet the archival-resolution requirements of preservation scanning.

However, filming first, and creating digital images by selectively scanning the film seems to be the least risky current preservation option provided that appropriate attention is paid to indexing the filmed collection.

Scanning first: If the choice is to create digital images as the first step in the preservation process the key decision revolves around the scanning resolution. Scanning original documents at a yet-to-be-determined "optimal archival" resolution means creating a balance that produces image quality comparable with photographic methods while minimizing the amount of data stored.

After scanning, image enhancement techniques are applied to improve image quality and the full high-resolution greyscale image is used to create high-quality film using an electron beam or digital computer output microfilm (COM) camera. The quality of the film created is governed by the scanning resolution and amount of greyscale data captured. (See **Appendix A.**)

At the same time, a parallel process uses the high-resolution greyscale image generated in the image enhancement process and converts it to a high-quality reduced resolution binary image suitable for information access. This very high-resolution image on film is the archival copy. The reduced resolution image in digital form can always be recreated from the film copy for only a few cents per page. Obsolescence is not a factor.

Timing and volume, two key factors: Of major concern when implementing a scan-first archival resolution preservation system is the amount of time that will elapse between image capture and conversion to film and the daily volume of documents being preserved. If the elapsed time is more than a day or so, and the volume is significant, it would be easier and less expensive to film first and convert to digital later. The length of time the archival resolution greyscale data has to be stored on magnetic or optical disk prior to filming, and the volume of pages to be captured, greatly affects the cost of the system. The longer this elapsed time and the higher the daily volume, the more attractive the film-first option becomes.

Simultaneous scanning and filming: At the 1992 AIIM show several vendors including Bell & Howell and Kodak introduced devices that allows simultaneous scanning and filming. These devices currently have low resolution (300 dpi) and are directed at the records management market, but they have potential for the preservation market at some future date. They both employ a very gentle belt feed that could accommodate all pages that have not yet begun to turn brittle. In addition, both have the capability to scan and film both sides of the page in a single straight through pass.

It should be noted that filming and scanning simultaneously has some of the same drawbacks as filming first. The film is created in exactly the same order that it was scanned; there is no easy way to build intelligence into the film; and if pages are skewed, misfed, or of poor quality they can only be spliced-in after the fact. *Scanning-first to get the page into digital form is the most flexible and efficient processing option for the future hybrid preservation system.*

Digital computer output microfilm (COM) camera: As data transmission and image enhancement technologies advance, and microprocessors become faster and more powerful, it will be cost effective to create intelligent digital film that is of higher quality than photographic produced film. It is this high-quality archival resolution digital film that could be the archival storage media for future preservation systems. The cameras capable of producing this film are: the Electron Beam Recorder from a company called Image Graphics, Inc., in Shelton Ct.; and a laser beam camera from iBASE Systems Corp., in Hayward, Ca. Both manufacturers claim that their camera can produce film that is comparable to photographically produced film.

This digitally created film can be intelligently indexed with blip marks, and bar codes to provide automated, accurate, and intelligent computer-assisted retrieval of specific pages or groups of pages from the film, thus providing a significant improvement in automated film access.

The additional intelligence that could be built into the film would allow computer-assisted monitoring programs to automatically migrate preserved documents between different levels and types of hierarchical storage consisting of magnetic disk, optical disc, digital audio tape (DAT), film, or other storage media in the most cost-effective manner. This system has the potential to eliminate one of the biggest costs associated with a large film archive: the cost of retrieving film to make copies. (Currently, at a film vault, that cost ranges from $15 - $30 per reel.) And because film is used as the system archive, any risk of obsolescence is eliminated. Optical disc would be used to provide storage for the higher-use data at levels of resolution that would satisfy the end-users information requirements (most likely 300 dpi binary).

Digital technology still under development: Some technology required to implement the hybrid preservation system, as defined herein, is still under development. High-speed, sheet-fed greyscale scanners, scanners that can scan bound books, high-speed binary and greyscale film scanners, high-capacity/high-speed reliable magnetic storage (parallel disk arrays), higher capacity write-once optical disks, high-speed greyscale digital COM cameras, and communications apparatus that can handle transmission rates of about 20 MB per second are all either unavailable or just becoming available. However, since digital imaging technology is in its infancy, these solutions will evolve rapidly. In fact, all will likely be available commercially within the next year or two.

Options for Converting from One Format to the Other

Hybrid systems must be designed to interface with past, present, and future technologies[11]. Although the design must anticipate these capabilities, operationally, the conversion can actually be accomplished through a preservation filming service bureau.

The migration path from past to present must allow preexisting microform (fiche and film) collections to be scanned and converted to a high-quality digital image format to improve access. This conversion process can take place almost automatically, depending upon the amount of intelligence built into the system and the film. It's simply a matter of mounting the right reel of film, spinning down to the correct frame and scanning the film, frame by frame. If intelligence has been built into the film during initial filming, then that intelligence can be used to index the images. The process is fast, efficient, and at a few cents a page, inexpensive. Microform scanners that support binary scanning at adequate access resolution exist. Archival resolution binary or greyscale film scanners are not yet available off the shelf, but should be in the near future. TDC, Mekel, and Photomatrix market both film and fiche scanners which provide greyscale output.

The migration path from present technologies to an older technology must allow the practitioner to create high-quality microfilm from archival resolution, greyscale digital images. This can be achieved by using a high-resolution electron beam or digital COM camera as previously mentioned. This process should be fast and efficient; but, depending on the resolution of the images, the cost of digital storage media, and the amount of time the digital data must be stored prior to creation of the film, the process may not be cheap.

The present-to-future migration path must anticipate storing not only binary and greyscale images, but also ASCII text, compound documents, audio, vector graphics[9], color images, and full-motion color video. All of these formats can be represented and stored digitally. Also, in the future, it will be necessary to provide the means to store an archival copy of materials that were never in print. For any data using the page metaphor, the system remains the same. The formatted digital data is composed into pages in memory and subsequently written to film using a digital COM camera. *Film is still the primary archive.*

Further discussion exceeds the bounds of this paper and is due to be covered in a future paper.

[9] A method for representing graphic drawings such as blueprints or circuit diagrams with mathematical formulas (instead of in raster or pictorial format).

ASCII Text and OCR

Extracting character code data from a page image is always an option: Technology currently available off the shelf allows pages in digital image format to be processed through an optical character reader to create ASCII text output. Certainly OCR could be helpful in automating the page creation, indexing, and abstracting process. The indexes and abstracts and/or full-text, stored in a separate database, combined with the proper automated system, could be used to gain access to the preserved information irrespective of format or storage media.

ASCII text--limited preservation usefulness: Character-coded databases are viewed as attractive because they require less storage space than image databases and are searchable. While this is indeed true, it is extremely difficult if not impossible to represent formulas, graphics, special characters, non-Roman languages, or pictorial data using just the ASCII character coded format; therefore, this technology is not directly applicable for preservation work. However, the ASCII text data could be combined with vector graphics and raster[10] imaging in a compound document format in order to recreate a replica of an original page thus solving the presentation problem. This would allow the researcher to search on the ASCII text, and recreate the original page with all of its graphics and halftones -- the best of both worlds. However, even if the chosen compound document format can meet all of the requirements for recreating a faithful reproduction of the original page, the storage media is still the critical part of the preservation equation.

The ASCII text or compound document format would be especially beneficially for books or other materials where most, if not all, of the information content is text. (See **Figure 7.**) A typical printed page of text-only data contains about 3,000 to 4,000 characters. Using the ASCII character-coded data format, one can represent any character in the Roman alphabet in one byte of data. Therefore, a text-only page can be stored in 3 to 4 KB. A digital copy of the publisher's original font set might also be stored as a file appended to the set of full-text ASCII pages. Assuming the output printer can handle the font set and print raster images, it could be possible to reprint--on demand--a facsimile copy of a book that looks very much like the original. Adobe has recently announced a product they call Carousel which is a font and platform independent Postscript[11].

Storing a page in a compound document format requires slightly less storage space and allows text data searching. The disadvantages are that it complicates the scanning process, sacrifices some of the editorial intelligence of the document, and requires more power at retrieval time to recreate the page. Line art or halftones on the page would be represented in

---

[10] A method for repoducing an image (on, for example, a display), where individual picture elements (pixels) within the image are addressd and represented in both the horizontal and vertical directions. These pixels can be turned on and off in the binary (black or white) mode, the greyscale (usually 8 bits per pixel) mode, or the color mode (usually 32 bits per pixel). Regular television pictures are created in raster format.

[11] A page description language developed by Adobe Systems. It is designed to translate text, line drawings and photographs created on a computer in conformance with its specifications into the proper bit-mapped dot pattern to recreate a page image on a screen or printer.

scanned image format and appended to the page. Given a scanning resolution of 300 dpi binary, and assuming that 50% of an 8.5 X 11 inch page[12] is halftone, the appended digital image file could be as large as 253 KB[13]. For comparison, a 600 dpi image satisfying the above constraints will be as large as 1.05 Megabytes (4 times as large because the resolution is doubled).

Fortunately, the typical journal being considered for preservation contains few halftones. For this example, let's say that the average page contains about 15% halftone content. Using the same formula as above, with 300 dpi resolution, but substituting the 15% factor (.15 for .5), and again assuming 2 to 1 compression, we can calculate the halftone content of this compound page at 79 KB. Adding 3 KB for text data, we can calculate the compound document size for this particular page (ASCII and image) at about 82 KB.

Since experience has shown that the average size of a journal size page with 15% halftone scanned at 300 dpi binary is about 100 KB, only 12 KB more than required to store the compound document, one must weigh the tradeoffs carefully before deciding to store pages in that format.

Other data formats: Many photographs and paintings can only be represented by the original or a very high-quality image. Other graphics can be represented in image format or vector format. The intrinsic value of a document is also a significant factor in determining the appropriate format for representation. Clearly, the Declaration of Independence, the Magna Carta, or the original Gutenberg Bible cannot be replaced by ASCII-coded data, but in image format they could retain much of their intrinsic value. Of course, for the researcher needing to see how the papers contained in these documents have aged, there is no substitute for the original.[12]

Image Access, Distribution, and Transmission

Access: The system should be structured to satisfy the users' information access needs while minimizing movement of large image files. Dedicated CD-ROMs could provide access to facsimiles of very high-use preserved documents in image format. Local collections of less frequently used documents could be stored in CD-ROM jukebox servers on local area networks (LANs). Film stored in small computer-assisted retrieval (CAR) systems could provide access to the least frequently used preservation materials. It is reasonable to assume that copies of other preserved documents would be stored in a similar way at other institutions or at a central site.[13]

---

[12] As mentioned earlier, a 5 X 9 inch typical book page is about half the size of the 8.5 X 11 inch page and therefore requires only about half as much storage space.

[13] 300 * 300 (8.5 * 11) / 8 * .5 = 526 KB divided by 2 for compression = 253 KB. We assume a compression ratio of only 2 to 1 because the high frequency black and white transitions present in all halftones do not compress well using CCITT run-length compression.

A user might be able to search any number of bibliographic catalogues from the desktop to identify specific materials that meet his/her research criteria. Making this database accessible over the Internet or some other network would allow widespread automated access to these treasures. The researcher could search for topics of interest or browse the image database(s) at the document structure or page level.[14]

Distribution: An average of 7,500 300-dpi compressed binary journal size page images fit on a single CD-ROM. This is equivalent to 50 books or 7.5 years of a journal publication. With production costs of about $0.50 per binary page image at adequate access resolution (including indexing and abstracting), mastering costs of $1,500, and unit costs of $2.00 per disc for 100 replications,[15] one can distribute the disc to 100 locations at a manufacturing cost of about $50.00 per copy. In the future preservation system, even if film is the archival media of choice, document images on CD-ROM discs could be the access and distribution vehicle.

When a request is received for a less frequently accessed document stored only on intelligent film, the film could be automatically located, advanced to the proper frame, scanned to create a digital image, and the image transmitted back to the requester. The digital copy would then be stored on optical disc. Subsequent requests for that publication could be serviced from the digital copy on optical disc. Once the document is stored on digital media it should remain there for some period of time (defined by the institution). If during that time, the document is not accessed then it is erased. Any future request for the document will be filled from the archival copy on film, and the process will repeat itself. This storage hierarchy is intelligently managed by a computer. The more frequently accessed preservation materials migrate to the faster, more expensive media, while infrequently used documents are migrated back to the slower, least expensive media.

Transmission: The National Research Educational Network (NREN) along with other commercial and non-commercial networks could allow widespread access to, and ordering and delivery of, preserved materials from various archives. Fax-delivered copies of preserved documents, could be ordered from other institutions or some central source. The requested documents could be retrieved, and if on film, scanned and converted to digital format and then fax-delivered back to the user within hours of request. High-speed networking along with digital imaging promises to make the knowledge of the ages available at the desktop.

COSTS

Handling Pages--Little Difference in the Cost: Michael Lesk, in an earlier report published in 1990 by the Commission on Preservation and Access entitled "Image Formats for Preservation and Access" (July 1990), concludes that microfilming a book costs about 10 to 15 cents per page. Digital image scanning was pegged at between 13 and 28 cents per page.[16] Our research indicates that current filming costs are slightly higher, and preservation imaging costs are about double those quoted by Lesk. The higher costs for preservation filming can probably be attributed to inflation and experience with the difficulties and corresponding costs of preservati filming. The higher costs for digital imaging can be attributed to the higher resolution scanning, and the high cost of storing these archival resolution images on optical disc.

However, the new generation of vacuum-fed, belt-driven, duplex scanners which have recently become available for handling non-brittle materials, along with the reduction in optical disc media costs, promises to reduce page imaging costs substantially. Some of these new scanners can capture both sides of a page in one second which is faster than any planetary camera; in addition, the newest ones can film the page simultaneously, using a planetary camera that is mounted on a camera stand above the feed belt. Of co se, platen scanners and through-the-lens scanners are also available for handling brittle materials in a very safe and efficient manner. These new developments guarantee that scanning costs should be no greater for digital imaging than for filming.

While page handling is one of the most costly function of preservation, one should not lose sight of the fact that the materials selection and acquisition process is also very expensive, so we want to make sure that whichever storage strategy is selected, the process does not have to be redone.

Cost of a Digital Image Preservation System

Regardless of which technology is chosen, the cost and technology necessary to implement and operate a preservation system is significant. For all but the larger institutions, these barriers could be insurmountable.

Note: The digital system implementation costs that follow have been increased by 50% from those presented in the reference (see endnote 17) to compensate for the fact that a preservation system must be implemented using archival resolution with technology available at time of publication.

It's interesting to note that when viewed on a per-page basis, the cost to implement the digital image systems described below range between \$0.15 - \$0.50, regardless of the size of the page. For further explanation of how resolution affects cost see **Appendix A.**

Digital system implementation costs: Digital image systems are usually configured to a certain capacity level:[14]

[14] Capacity references are editorial comments by the author.

Stand-alone microcomputer-based systems: This system is capable of non-critical workloads up to 300,000 pages per year. Some 47 percent are priced at less than $60,000, another 40 percent between $60,000 and $150,000. ($0.24 - $0.50 per page)

Networked microcomputer-based systems: Depending on design, these systems are capable of critical workloads of between 150,000 and one million pages per year. Some 23 percent cost between $60,000 and $150,000, 70 percent over $150,000. ($0.20 - $0.40 per page)

Minicomputer [or microprocessor]-based systems: Capable of workloads from one million to five million pages per year. Sixty-nine percent cost under $450,000, 27 percent range from $450,000 to $750,000. ($0.15 - $0.45 per page)

Mainframe [or multi-processor]-based systems: Designed to handle workloads of over three million pages per year. Forty percent cost under $450,000, 33 percent between $450,000 and $750,000, and 27 percent over $750,000.[17] ($0.15 - 0.25 per page)

The components of a typical digital image system are listed in **Figure 8.**

Digital system operating costs: The cost of creating a digital page image, indexing it, and storing it on optical disc on a custom-designed in-house system is between $0.30 and $1.20 per page, depending on volume, size, type of documents, condition, amount of halftone content, amount and type of indexing, resolution, and amount of image processing required. Capturing a binary 300-dpi image of a good-quality text page, compressing it, and storing it on optical disc with simple indexing can be done for between $0.30 and 0.55 per page. On the other hand, capturing an archival resolution image with greyscale, complicated indexing, and image enhancement will cost between $0.50 and $1.20 per page. (Estimated prices originate from actual experience modified by an informal survey of image processing sites by the author.)

Contract preservation imaging costs: Contract preservation imaging is estimated to cost between $0.50 and $2.50[15] per page. Capturing a binary 300-dpi image of a good-quality text page, compressing it, and storing it on optical disc with simple indexing can be done for between $0.50 and 1.25 per page. On the other hand, capturing an archival resolution image with greyscale, complicated indexing, and image enhancement will cost between $1.00 and $2.50 per page. The above costs include indexing and storage on optical disc. (Estimated prices originate from actual experience modified by an informal survey of image processing sites by the author.) These costs may seem quite a bit more costly than in-house scanning; however, if all direct and indirect costs are included, and intangibles are factored in, contract scanning would probably be found comparable. It should be noted that image service bureaus can provide expertise, guaranteed workmanship, liability, and diverse

[15] Estimates are used here because service bureaus have had very little experience with preservation scanning. Estimates were arrived at by surveying service bureaus.

equipment relieving the burden on the institution to provide these facilities or hire and train staff. However, the selection of the imaging service bureau must be done very carefully since most are unfamiliar with the quality requirements necessitated by preservation processing and tend to underestimate the costs involved. Preservation imaging should also be more costly than preservation filming.

Optical disc drives and media costs: The typical 12-inch optical disc player costs $16,000 and uses double-sided discs that have total capacity of about 4 GB each and cost about $300 when purchased in quantities.[18] The cost to store an archival resolution book size page on a 12-inch disc is about $0.085 per page (media only).

The typical 5 1/4-inch optical-disc player costs $3,000 and uses a double-sided disc with a total capacity of about 600 MB and an average cost (purchased in quantities) of $100.[19] The cost to store an archival resolution book size page on 5 1/4-inch disc is about $0.19 per page (media only).

These costs double to $0.17 and $0.38 respectively for journal size pages.

Cost of a Micrographics System

Expected workload: As with digital systems, micrographics-based preservation systems can be configured based on expected workload. An experienced operator can film about 200 exposures per hour (2 pages per exposure). This works out to about 9 seconds per page, 3,000 pages per 7.5 hr shift, or about 750,000 pages per year, per operator. The Cornell/Xerox project has achieved scanning rates (600 dpi binary) of over 1,500 images per day for three weeks.[16] This is about half the rate achievable for film operators; however, it includes some indexing and QC. At a fully loaded labor cost of $12.00/hour, the filming costs work out to about 3 cents per page, with another 1.5 cents for QC. Add in system depreciation, film costs, duplication costs, packaging and labeling, retakes, storage, handling, insurance, facilities overhead and profit, and we arrive at a cost of about 15 cents/page for filming the best materials. Filming old or brittle materials could easily double the cost.

Micrographics system implementation costs: Naturally, different sizes of micrographic preservation systems are required for different preservation projects. It's interesting to note that when viewed on a per-page basis the cost to implement the micrographics system described below range between $0.04 - $0.35. These are about 1/3 less than the costs to implement a digital image preservation system of comparable capacity, and are based on purchasing refurbished cameras.

---

[16] Kenney, A.R. & Personius, L.K., Update on Digital Technologies, Newsletter Insert, Commission on Preservation and Access, Nov. - Dec. 1991, Pages 1-6.

A one-camera system, low-speed processor system: This system is capable of non-critical workloads of up to 500,000 pages per year. Costs are between $70,000 and $90,000. ($ 0.14 - $0.18 per page) See **Figure 9a.**

A multiple-camera, low-speed processor system: Depending on design, capable of critical workloads of between 650,000 and three million pages per year. Costs are between $150,000 and $250,000. ($0.08 - $0.38 per page) See **Figure 9b.**

A multiple-camera, medium-speed processor system: Capable of workloads from two million to five million pages per year. Costs are between $250,000 and $400,000. ($0.05 - $0.20 per page) See **Figure 9c.**

A multiple-camera, high-speed processor system: Designed to handle workloads over four million pages per year. Costs are between $400,000 and $800,000. ($0.10 - $0.20 per page) See **Figure 9d.**

NOTE: Large-scale film processing operations require air conditioning and humidity control, chemical holding, storage, disposal facilities, and silver recovery facilities.

Micrographics system operating costs: The cost to perform in-house preservation filming is estimated at approximately $0.10 to $0.18 per page[17]. However, it is doubtful that these costs include any indirect or overhead components. Also, in order to operate an in-house facility the institution must deal with the following issues: 1) air conditioning and humidity control, 2) building a darkroom to house the processor and for handling film, 3) plumbing for the processor, 4) designating a secure storage space with a controlled environment for storing the camera negatives, 5) accumulating the necessary test equipment (both chemical and photographic) needed to create high-quality film, and 6) hiring a photographic technician or engineer to run the operation.

Contract microfilming costs: As with contract scanning, contract preservation microfilming may appear at first glance to be more expensive than in-house filming; however, in actual fact, if all the in-house costs were accounted for, preservation microfilming would be found comparable. In addition, a service bureau can provide expertise and advice, material preparation, liability, processing, bibliographic services, and diversity of equipment, among others. The cost of creating microfilm at a service bureau is between $0.07 and $0.50 per page. Micrographics service bureaus charge between $0.07 to $0.15, averaging $0.08 per page to create 16mm standard document storage film. On the other hand, preservation microfilming vendors charge between $0.10 to $0.50 or more, averaging about $0.15 per page to create archival microfilm.[18] The distinction here is important. Preservation

[17] Source: Survey of other preservation micrographic sites by author.

[18] Source: informal survey of preservation microfilming service bureaus by author.

microfilming is more costly because of the higher standards and more stringent processing requirements. Preservation microfilming costs include creation of the master and two copies, quality controlled, labeled, and packaged. Polysulfide treatment can be added for about $3.50 per reel ($9.00 for all three copies). This does not include any automated indexing. Without indexing, of course, access is restricted.

Film storage and duplication costs: 35mm camera negative silver gelatin film costs about $0.10 - $0.12 per foot; silver print film costs about half that amount. Given that on the average one can store 12 frames per foot with two exposures per frame; and if the original camera negative is stored in a vault as the archival copy and two other copies are made--one for reprinting and one for the end-user--the film cost to preserve a page (making all three copies as described above) is about $0.01 (cost of film only). Additional copies can be made on silver duplicating film quickly and at relatively low cost using a roll-to-roll contact microfilm printer ($15 per reel at time of filming, double that later). Silver film is a requirement for preservation filming.

Preservation Cost Summary: Currently, film is the most economical technology for preservation (**see Appendix D**). However, the micrographic based preservation system is expected to become more expensive to operate over time simply because it is labor intensive and the cost of labor will continue to increase, while advances in micrographic technology will not increase productivity enough to offset these increasing labor costs. On the other hand, for the digital preservation systems, productivity increases will result from rapid technology advances, which are expected to accelerate rapidly over the next several years.

The practitioner should therefore become familiar with digital technology and begin planning for its use. Currently, the best use of imaging technology for preservation is to provide selective access capabilities at adequate binary resolution to the preserved collection. High-resolution, archival-quality, greyscale scanning is still expensive. It will be about another year or two before a combination of decreasing prices and advances in computer and imaging technology will make an archival resolution image preservation system cost-effective.

Finally, the institution could decide to follow the approach pioneered by Cornell, which is to create a high-quality copy on acid-free paper from scanned data at 600 dpi binary. The idea is to create a permanent, not archival, copy that can go back on the shelf. A copy could also be kept for archival processing sometime in the future. The Cornell results show quality of the output document equivalent to or in some cases, better than the original. The solution is inexpensive, practical, and effective.

Conclusion: The requirement for high-resolution greyscale imaging and the cost of optical disc storage is a major reason why archival preservation using imaging technology is still substantially more costly than archival filming. Film is the least costly storage media. A 125-foot roll of film, created to RLG preservation specifications, can contain about 2,700 nine inch pages at a reduction ratio of 12X with two pages per frame. The film cost is approximately $15.00. The same number of archival resolution images (9 X 5 inch page @ 600 dpi with 8 bits/pixel compressed 15:1) would require 3.0 GB of storage space. That's the equivalent of about one Write-Once Optical Disc at a cost of about $300.00. In this particular example, optical disc storage costs are 20 times more expensive than film. However, advances in imaging technology, cost reductions in the digital storage costs, and increasing costs for preservation filming argue for using film as necessary to satisfy critical needs, but beginning the switch to the hybrid digital image preservation system as soon as technically feasible. In fact, if the objectives of the system do not require high-resolution greyscale scanning (i.e., very few halftone pictures, as is the case for the materials being preserved in the Cornell Project), and the 600-dpi resolution offered by, for example, the Xerox DocuTech system, is considered adequate, then the practitioner will probably find digital imaging equivalent to filming in cost.

RECOMMENDATIONS

Get Involved: The preservation manager should feel comfortable in joining with the technical experts, side by side, to develop the science. There are key questions to be answered and standards to be formulated. We must become proactive, recognizing that good preservation systems will only be developed when the preservation community takes an active role in the development process. We can build alliances with digital image vendors and information suppliers. We can educate the developers about preservation requirements and in turn, be educated about the technology. We can work with the technical experts to develop strong requirements and specification documents. We can set the tone for how these systems will evolve.

Understand the Technology: Preservation has developed quickly as a science, but some basic questions remain unanswered. Preservationists must weigh a variety of concerns when choosing a preservation format. In the parallel universes of micrographics and digital imaging, this is no easy task. Digital imaging is as misunderstood for preservation work as micrographics is commonplace. For instance, what is the minimum digital image resolution and greyscale combination that will satisfy the archival requirement for preservation? At what point will digital resolution be equivalent to film resolution? Does it need to be, or should the standard be changed to consider low-contrast page areas more? How can we influence vendors to develop the kind of high-resolution scanners, book scanners, high-bandwidth communications, etc., that digital image preservation work requires? Also, the matter of electronic media obsolescence and how it applies to archival storage is not well understood or generally accepted in terms of preservation economics or policies. Finally, the access, transmission and distribution requirements must be understood and evaluated, and their economic impact factored into the equation.

Minimize Risk: In the world of information science, technology travels faster than the speed of decision-making. Adopting an electronic publishing preservation strategy requires a tremendous investment of resources. A backlog of several billion pages awaits conversion now. Brittle research materials are deteriorating rapidly. And although a hybrid system is within sight, the vanishing documents will not wait. To minimize risk, a solution that uses today's micrographics technology can and should be implemented, but this solution must anticipate the evolution of imaging technology. Preservationists should be aware of future access needs and consider the best methods for filming material for later conversion to digital formats. There is no doubt that digital imaging will play a large part in the future of preservation science.

Prepare for the Future

If the technology were available: Fc. anyone considering a preservation imaging system, the design and implementation will likely take 12 to 18 months. By that time much of the required technology should be available. However, there is no reason to stand by and wait for technology to advance while delicate preservation materials continue to deteriorate. The important thing is to preserve materials in a recognized archival media for future generations. Film is currently recognized as that medium. As long as the film created is of high quality, and has a good low-contrast rendering of the halftones as well as a high-contrast rendering of the text, it can be scanned to digital when the complete preservation system is implemented.

The future digital solution: An effective preservation system should be designed so that the material is scanned at the optimal archival resolution with eight bits of greyscale per pixel. This high-resolution data will be further processed (as defined by the objectives for the project) using mathematical image enhancement filters, and finally be written to film to create an archival image that can always be accessed. A parallel process will convert the input data to a high-quality reduced resolution (adequate access resolution), enhanced, binary image that will be written to optical disc, which would guarantee improved access and excellent end-user print quality.

The long-range system: Twenty years from now we're likely to see high-quality color page images stored using laser holography in a diamond composite storage medium that will cost less than one-tenth of a cent per page and last virtually forever. Compression algorithms will recreate pages from less than five percent of the data, and transmission costs will be 1/20th of current costs. The storage medium will be self-contained with built-in intelligence (the processor and the memory will be one), it will have the capability to monitor itself, correcting faults automatically, and when its error rates are projecting end-of-life, it will have the capability to schedule that it be rewritten. Since it has a built-in processor, it could also contain all the necessary software to recreate electronic page representation back into eye-readable form regardless of storage format. The question of obsolescence will become irrelevant. Systems will automatically monitor user access, read errors and storage costs of page images, and automatically migrate pages throughout the storage hierarchy depending on preprogrammed factors. Such a system will manage and preserve all digital materials automatically. If this seems far-fetched, remember that the PC is only about 10 years old.

An optimal mix: Today, according to a traditional definition film is the only truly archival medium. It will not become obsolete in the foreseeable future. Optical disc will be viewed as the permanent, low-cost, removable, random access storage media. Magnetic products (tape and disk) will continue to increase in storage capacity and reliability while decreasing in cost. Magnetic disk will provide temporary working storage for all work-in-process on all future image systems. Optical tape, too, bears watching. In configuring the ideal image storage system, the knowledgeable designer will construct a hierarchy of storage that takes advantage of the strengths, access characteristics, longevity, and cost of each storage product to produce the greatest benefit at the least cost.

APPENDIX

Appendix A

RESOLUTION--A KEY DESIGN PARAMETER

The Single Most Important Factor: Resolution is the single most important factor to consider when designing the digital image preservation system. It is critical to have an in-depth understanding of both film and digital image resolution to make informed design tradeoffs involving quality versus cost.

Designing around the amount of data: A 600-dpi image is composed of 360,000 pixels per square inch, which amounts to four times the 90,000 pixels per square inch areal density generated at 300 dpi. An 8.5 x 11 inch page in the binary format at 600 dpi resolution requires 4.207 MB of storage before compression. To get an idea of just how much data this is, it would take almost four hours to transmit one 600-dpi binary image at 2400 baud, or, it would take three double-density floppy disks to store the image. If we want a truly archival image with greyscale, multiply this figure by eight. It's the amount of data being captured, moved, and stored that most affects the design of the preservation system.

How high resolution affects cost: A high-resolution archival system must be designed with more powerful processors, higher capacity communication channels, more random access memory (RAM), more magnetic disk storage capacity, more scanners, and possibly custom hardware to handle the compression, decompression, and processing of high-resolution images. These components increase the cost of the system. In fact, the capital costs and operating costs of a digital image preservation system are directly proportional to the resolution it is designed to handle. The resolution decision is absolutely crucial. Not only does it affect the design of the system, but it also determines the maximum possible quality for each image captured.

How Much Image Resolution Is Required? If the system's objective is to preserve the information content of the page, it can be accomplished at a much lower scanning resolution than would be required to preserve a high-fidelity copy of the original. A resolution of 300 dpi (adequate access resolution) will preserve about 99.9 percent of the page's information content. One need only examine a page transmitted by a typical fax machine to see that even at a resolution of 200 dpi (which is the resolution of today's fax--see **Figure 10**), all but the smallest type fonts and finest lines are faithfully, albeit crudely, rendered. Most loss occurs in the area of the halftones, yet most of the intelligence in the page is preserved.

Capturing small type sizes: Let's assume one design requirement is to be able to read footnotes from the captured page that are in four-point type. A four-point character has a height of 4/72 of an inch (each point is 1/72 of an inch). Assuming that the theoretical character is formed within a cell that is five lines high by five lines wide, each line that forms the character would be approximately 1/5 of the character's total height, or in the case of a four-point character, the line width is 4/72 * 1/5 = 1/90 or .011 inch wide. To capture this character legibly, the scan resolution must be at least fine enough to have two or more scan lines (assume three) overlay each line that composes the character. This means that each scan line must be no greater than .0037 (.011/3) inch, which translates into a scanning resolution of about 300 dpi (1/.0037 = 270).

Figure 11a shows an enlarged portion of the IEEE test chart scanned at 300 dpi. (The IEEE chart is shown in its entirety in **Figure 4.**) Looking closely at the four-point type roughly in the center of the page, we can see that there are approximately 12 scan lines from the top to the bottom of the character. (The black lines are electronically generated every eighth line.) This measurement is designated the "X-height" of the character (see **Figure 11b**). The actual "body height" of the character is approximately 30% larger. We can calculate using the formula 4/72 * 300 dpi = 16.6 that approximately 16 scan lines cross the body height, which is also known as the point size. (Type size expressed in points does not refer to the actual dimensions of the character but to the height of the metal surface on which the raised design is produced for typesetting.)[20] From the close-up in

Figure 11a, we can clearly see that four-point type is at the limits of the resolving capability of 300 dpi scanning. The two-point type below and to the right in **Figure 11a** has been completely lost at 300 dpi.

At 400 dpi (see **Figure 12**) the four-point type, located at the bottom of the page and slightly to the right, is much more legible. **Figure 11a** and **Figure 12**, graphically illustrate the role resolution plays in capturing a high fidelity replica of the original page.

The typeset parameters for differing groups of documents vary according to font type, document age, background coloration, and other factors. To verify the above calculations for a specific set of documents, the designer should test a random sample of the material to be scanned, along with some resolution test targets, to determine the smallest fonts that must be captured. The scanning system should provide the capability to zoom in on (magnify) a small portion of the scanned page (as we have done in **Figures 11a and 12**), even down to the individual character level, to determine if the characters are being properly formed. The pages should also be viewed on a full-page display (100 dpi) to verify that the reduced resolution display will be adequate for operator processing and quality control.

In addition, a greyscale test chart should be used to measure the number of levels of grey being reproduced.

Finally, the test pages should be printed to verify that the print resolution is sufficient. If halftone fidelity is important, the appropriate image enhancement processes might be required as part of the system.

Halftone resolution: Surprisingly, it takes less input resolution to produce good digital halftones than to render good quality text. A simple formula for input is: a scanning resolution of about 1.5 times the desired output screen ruling[19] is sufficient for scanning. To get a good 133-line screen (equivalent to the resolution in a typical magazine), images need only be scanned at a minimum resolution of 200 dpi with greyscale. This rule works as long as the image is printed at the same size as it was originally scanned.

On the output side: the relationship between printing resolution (line screen) and number of greyscales can be determined by the following equation: number of greylevels = (printer output resolution / line screen) squared + 1. If you try to print the same 133-line screen on a 300-dpi printer the result is 6 levels of gray. However, if you drop the line screen down to 50 using the same printer you get 37 greylevels [(300/50)squared + 1 = 37], which is about optimal for a 300 dpi laser printer. Of course, a 50-lpi diagonal screen is coarse, but as you can see from **Figure 13**, it renders the halftone with some degree of fidelity.[21] If the resolution of the output device is held constant, then as screen resolution increases, the number of greyscales decreases. This is why a high resolution output device is necessary to render high screen resolution while at the same time reproducing a high number of greylevels.

Archival resolution: As stated above, binary resolution on the order of 1,000 dpi is required to create an image comparable in quality to an image stored on film. Theoretically, one should operate at resolutions as close to this as possible. But due to the high cost of doing so, it is just not practical with today's technology. If the objective is to capture every detail of the smallest type fonts, the finest of the graphic art lines, and produce an accurate rendition of any halftones on the page, the required resolution would have to be about 600 dpi or more. Archival resolution is designed to preserve a faithful replica of each document.

---

[19] Screen ruling, as used here, is defined as the distance between the halftone cells measured on an angle and from the center of each cell. The angle is about 45°. In this case, for example, a 7 X 7 cell would have a screen ruling of (square root of [7 squared + 7 squared]) = 9.9, equivalent to a 30 line screen; therefore the number of grey levels for this cell would be 300/30 squared + 1 = 101.

Greyscale scanning can improve page quality: Regardless of the resolution, digital page image quality can be improved by using scanners that capture the image in shades of gray. The additional greyscale data can be processed electronically to sharpen edges, fill in characters, remove extraneous dirt, remove unwanted page stains or discoloration, and, in effect, create a much higher quality image than possible with binary scanning alone. A major drawback to scanning in greyscale is the large amount of data captured. Since this is the case, methods must be found that will take advantage of the additional greyscale input data to produce a high-quality image while minimizing the amount of data needed to be stored. Image enhancement is just such a method.

Image enhancement: After scanning the page with a greyscale scanner, the digital greyscale data can be mathematically manipulated to automatically decompose the page into text/ line-art areas and halftone areas and to process each area of the page with mathematical formulas (filters) that maximize content quality on each area of the page separately. For example, on the text area of the page, an edge detection filter could more clearly define the character edges, a second filter could remove the high-frequency noise (stray ink spots or dirt), and, finally, another filter could fill in characters. Greyscale areas of the page could be processed with different filters to maximize the quality of the halftone. See **Figure 14.**

In addition, the contrast range of digital image greyscale data can be increased--that is, the various greylevels captured during the scan can be recorded into a histogram with values from zero to 256 (if scanning at eight bits per pixel). As an example, let's look at an original photograph, a reproduction made on a typical photocopier, an image scanned in binary mode, and a mathematically enhanced image. (See **Figure 15a.**) It is clear that the enhanced image comes much closer to the quality of the original than any of the other reproductions. Next we'll view the same page after greyscale scanning, mathematical enhancement, and rehistogramming. The image at the top of **Figure 15b** and the graph show that most of the values captured in the original histogram are spread over a fairly narrow range of the greyscale, from zero to 100. It is easy to spread those sample points over the entire greyscale range to improve contrast and make unreadable areas of the page readable. This is done by remapping a narrow range of greyscale onto a wider range and separating the levels so there are about 30 gradations in the image area. (See the two images on the lower half of **Figure 15b.**) This rehistogramming technique, which is also called "stretching the gamma," very effectively increases the quality of an image[22].

Finally, stains and discoloration can be removed using background filters, and the page can be restored to look much as it did when originally published. After some processing, the enhanced page image can be written to film as a greyscale image or it can be "thresholded" to remove greyscale and reinterpret the data on optical disc as an enhanced binary image.

Standard compression algorithms: One way to reduce digital image page storage requirements is to compress data. Binary data compression is accomplished by algorithms known as the CCITT Group III and IV Facsimile Compression Algorithms[20]. They work by removing redundancy. The algorithms represent strings of either black or white pixels (run lengths) by a code. These codes are a shorthand way of representing the black ink and white space on the page. Facsimile algorithms are **lossless** and completely reversible--that is, the original scanned image can be re-created exactly from the compressed data. Average compression ratios of ten or 20 to one are possible, which means that an exact replica of the page can be recreated from as little as five percent of the scanned data.

The greyscale compression algorithms developed by the Joint Photographic Expert Group (JPEG) promise high-compression ratios for greyscale data. This compression works by finding areas of the page that have some common tone, shade, color, or other characteristics and representing this area by a code. But this compression is achieved at the cost of some loss of data. Preliminary testing indicates that a compression of about ten or 15

[20] See M. Stuart Lynn's glossary, page 47.

to one can be achieved without visible degradation in image quality. Since this algorithm is not completely reversible, more testing must be done before it can be used with complete confidence.

Equating scanner resolution and film resolution: Film resolution is measured in line-pairs per millimeter. By definition, one line-pair is equivalent to two digital image scan lines. To scan an original page at 600 dpi with the objective of storing that page on 16mm film at a reduction ratio of 24X, resolution could be compared as follows:

Since one inch equals about 25 millimeters and since the reduction ratio being used is 24X, then one square inch on the original will be recorded on approximately one square millimeter of the film. Given the above, a scanning resolution of 600 dpi (300 line-pairs/inch) translates into a film resolution of 300 line-pairs per millimeter. However, because of the Nyquist sampling error (see below), one third of this resolution could be lost. Therefore, for this example, the effective resolution on the film is about 200 line-pairs per millimeter. When working with a reduction ratio of 24X, the following simple rule can be used (see **Figure 16**):

film resolution (line-pairs per mm) = binary scanning resolution (dots per inch) / 3

For 35mm film with a reduction ratio of 12X, where approximately one inch of the original page maps onto about two millimeters of the film, the resulting film image is about twice as large as the image generated in the example above. The simplified rule for 12X reduction could be stated as:

film resolution (line-pairs per mm) = binary scanning resolution (dots per inch) / 6

It should be noted that this formula includes sufficient resolution to overcome the sampling error.

The general formula is:

film resolution (line-pairs per mm) =
(binary input scanning resolution (dpi) / 2 scan lines per line-pair) * (reduction ratio/ 25.7 mm per inch) * .66 Nyquist error

The Nyquist sampling theorem: line-pairs scanned with equivalent-sized pixels have an equal probability of coming out black or white since the scan lines do not line up precisely with the black lines in the image. Therefore, a reduction in the pixel size, which is the same as doubling resolution, is needed to ensure accurate capture of image detail. This sampling error phenomenon is known as the Nyquist sampling error. [23]

Digital Image Scanners: How They Work

Binary scanners: Most scanners available today operate by moving a light-sensitive CCD array down the page at a fixed rate. The CCD has a sufficient number of discrete sensors (CCD elements) to generate the specified number of samples per inch (resolution) in the horizontal direction, multiplied by the width of the page. For example, to sample an 8.5 inch wide page at 300 dpi, the CCD array would require a minimum of 2,550 elements. See **Figure 17.**

The speed of the electronics combined with the rate at which the array moves down the page governs the vertical resolution. Each CCD element records the amount of light reflected off the page as measured by the changes in their electrical charges, like a sort of thermometer. This CCD thermometer records a value, between zero and some upper limit for each dot (pixel). In binary scanners a threshold value is selected to convert this **analog** representation of light into a binary value of either black (0) or white (1). We can compare the threshold value to the freezing point (0^0 Centigrade) of water. On a typical Centigrade thermometer, if the temperature is above 0^0 C, water will not freeze. If the temperature is below this level, water will freeze. In the scanner, everything below the selected threshold point is defined as black (0) and everything above that point is white (1). In binary scanning, no greylevels are preserved.

Binary page storage requirements: To determine the uncompressed size of a journal page stored at Adequate Access Resolution, the formula is:

$$BS = L * R * W * R; \text{ where}$$

BS is binary page storage requirement (bits)
L represents page length (in.)
W represents page width (in.)
R represents the scanning resolution (dpi)

Given an 8.5 x 11 inch page and a scanning resolution of 300 dpi, the above formula gives an uncompressed storage space of 8,415,000 bits or, dividing by eight, 1,051,875 bytes. If we assume a compression ratio of 12:1, which is typical for CCITT Group IV compression of an average journal page, the per-page storage requirements can be reduced to an average of about 90 kilobytes (KB). Book pages can be stored in about 45 KB due to their smaller size and general lack of greyscale.

Greyscale scanners: Higher quality scanners can scan greyscale--that is, they have the capability to represent the amount of light being reflected off the page at each pixel by a value recorded by the CCD element. To return to the thermometer analogy, we are now interested in storing the exact temperature represented by the reading on the CCD thermometer, not just whether the temperature is above or below freezing. The number of greylevels recorded determines the number of bits required to store each pixel. Sixteen greylevels requires four bits (two to the fourth power) to represent it. At eight bits per pixel, the scanner can represent up to 256 levels of gray. The eight bits per pixel metric is the level usually referred to when discussing high quality monochrome scanning requirements because the eight bits will allow 256 greylevels to be stored. Although studies indicate that the average person can only perceive about [illegible]0 levels of gray, capturing 256 levels provides sufficient over- sampling of the data to reconstruct at least 32 discrete greylevels.

Greyscale page storage requirement: Since in the greyscale image we are storing eight bits of data for each pixel sampled, the formulas given above for binary storage space must be multiplied by eight to give the formula for per-page greyscale storage space:

$$GSS = 8(L * R * W * R); \text{ where}$$

GSS is greyscale storage space requirement (bits)
L represents page length (in.)
W represents page width (in.)
R represents the scanning resolution (dpi)
8 is the number of bits per pixel or depth of greyscale

Capturing the same 8.5 x 11 inch page in greyscale at a scanning resolution of 600 dpi and a depth of eight bits per pixel, the above formula gives an uncompressed storage space of 269,280,000 bits, or dividing by 8 = 33,660,000 bytes. Assuming a JPEG compression ratio of 15:1, which is the maximum attainable without perceptive loss of data, the average journal page captured at Archival Resolution requires compressed greyscale storage space of 2.244 Megabytes, 1.13 Mbytes for a book size page.

Printing

The laser printer: The laser printer is currently the primary output engine for digital image printing. The laser printer is used almost exclusively because interface boards are available that connect to the laser engine directly and can drive it at video rates (about one megabit per sec.), dramatically increasing print speeds. Printing a digital image through a serial or parallel printer port would take about a minute per page. Printing through the video interface takes about eight seconds a page. Other arguments for the laser printer include their high-resolution printing capability, size, convenience, and quiet operation.

Creating a halftone: In standard printing processes, pictures are formed by placing ink dots in a pattern that creates the illusion of a photograph. The resulting image is called a halftone. The thickness and spacing between the dots are constant, but the dot size varies. The screen ruling measures the frequency of halftone dots at an angle. A newspaper has a screen ruling of about 80 dpi; a medium-quality magazine about 133 dpi; and a high-quality art book might have a screen ruling of 150 to 160 dpi. Halftone dots that are closer together tend to look more like original photographs.[24]

Halftone printing with a laser printer: The halftone printing dot is different from the greyscale scanning dot. The dot created by a greyscale scanner contains greyscale information (depth) that represents the degree of light (shade of gray) reflected from the page at that particular point on the page. However, since printers can only print black dots, a halftone printer dot is actually a group of black dots arranged in a cell that gives the illusion of a halftone.

A key objective of any imaging system is to reproduce a high- quality, high-fidelity rendition of each page image. A laser printer has a difficult time representing halftones because it synthesizes greyshades by grouping black dots together into grids or cells (sort of a super pixel) that represent the halftone dot. These cells containing certain patterns of black dots are interpreted by the eye as a halftone. (See **Figures 18 and 19.**) For a 300-dpi laser printer, the optimal screen ruling has been determined by testing to be about 50 of these cells (halftone dots) per inch. This gives the right balance between the coarseness of the screen pattern and the amount of greylevels this group of patterns can represent.

Finally, PC-boards are available that use techniques for modulating the printer's laser beam to create smaller dots at more frequent intervals, thus increasing the horizontal resolution and consequently increasing the number of levels of greyscale that can be produced on the standard 300-dpi laser printer.[26]

Appendix B

A SUMMARY OF STORAGE POSSIBILITIES

Film
o low-cost archival storage
o should experience a rebirth due to use in digital-imaging systems
o core technology for archival storage on imaging systems for at least the next several years

Magnetic Disk:
o high-speed random access
o will continue to be used for high-speed buffer storage and temporary working storage on fileservers and workstations in digital imaging systems

Magnetic Reel Tapes:
o slow sequential access, low cost
o will become extinct in five to ten years

Optical Disc:
o random access, removable, medium speed
o will be the core data storage technology for providing low-cost random access in imaging systems during the 1990s and beyond
o archival issue will be solved, obsolescence will require recopying

CD-ROM (660 MB, read-only):
o stores approximately 330,000 character-coded text-only pages
o 6,000 to 10,000 300 dpi compressed images
o ideal distribution and database publishing medium
o increase in capacity and throughput due shortly

Optical Card:
o ten MB of laser-written data on credit-card-size card
o important medium for notebook PCs

Helical Scan Tape (new technology, shows promise for back-up and possible distribution of large data files, including image files):
o 4mm digital audio tape (DAT) at 1.2 to 2.4 GB
o 8mm video at 2-5 GB
o both have robotic handling systems available

New Technology Optical Tape:
o experimental technology, first deliveries in '91
o single 12" optical tape stores the equivalent of 1,500 CD-ROMs or one terabyte[21] of data
o cheaper than any other form of storage; may compete with film for storage of greyscale images in future
o 28 second average and 60 seconds maximum end-to-end access time claimed

---

[21] One terabyte is one trillion bytes or equal to 1,000,000 megabytes.

Appendix C

DATA STORAGE COSTS[22]
(media only)

Cost per megabyte (in U.S. dollars)

non-removable hard disc	$15.00
removable hard disc	$ 6.00
CD-ROM[23]	$ 2.27
magnetic tape	$ 0.30
microfilm	$ 0.10
optical disc	$ 0.08
paper	$ 0.07
8mm video tape	$ 0.006
optical tape	$ 0.005

[22] Media costs only, or equivalent media costs to store about 10 fairly complex 600-dpi binary image pages @ 100 Kbytes each.

[23] Assumes that the CD-ROM is used for preservation purposes only and therefore, only one disc is created. Cost of mastering 1,500 / 660 MB / = $2.27 per MB. Cost per disc is much lower when CD-ROM is a distribution medium and numerous copies are produced. At 100 copies, the cost is reduced to about $0.02 per MB.

Appendix D

Preservation Cost Summary

(All costs are per page)

Category	Film	Digital
Per-page access cost (Lesk)	$ 0.10 - 0.15	$ 0.13 -0.28
System Implementation cost	$ 0.04 - 0.35	$ 0.15 - 0.50
Operating Cost	$ 0.10 - 0.20	$ 0.30 - 1.20
- Adequate Access Resolution		$ 0.30 - 0.55
- Archival Resolution		$ 0.50 - 1.20
Contract Preservation Costs	$ 0.10 - 0.50	$ 0.50 - 2.50
- Adequate Access Resolution		$ 0.50 - 1.25
- Archival Resolution		$ 1.00 - 2.50
Media Cost (book-size page)	$ 0.01	$ 0.085
Backup Cost (book-size page)	$ 0.005	$ 0.085

Appendix E

STANDARDS

Anyone contemplating preservation conversion should be aware of the numerous standards that apply. These include standards for film, scanning, compression, optical discs, and computers. Specific standards exceed the scope of this paper. However, the reader is encouraged to contact the following.[27, 28]

1. **Optical disc:** International Standards Organization (ISO), particularly Sub-committee 23 of TC97, (Joint Technical Committee- -JTC1) and TC171, the International Micrographics Standards body, for standards covering optical disc. Also, TC42 for photographic technology.

2. **Scanner test targets:** Association for Information and Image Management (AIIM), particularly C-13.1 committee for scanner test targets.

3. **Various digital image standards groups:** Other standards-making or influencing groups include: Association for Information and Image Management (AIIM), National Institute of Science & Technology (NIST), National Information Standards Organization (NISO), American National Standards Institute (ANSI), Special Interest Group on CD-ROM Application and Technology (SIGCAT), Digital Image Applications Group (DIAG), Federal Council on Computer Storage Standards and Technology (FCCSSAT), Optical Digital Data Disks sub-committee of Accredited Standards Committee X3 (TCX3BII) (3). Two important standards are ANSI X3B9 and X3B11 for re-writable and write-once optical discs, respectively.

4. **Compression:** CCITT (Comité Consultative Internationale pour la Téléphonie et la Télégraphie) for facsimile compression standards

5. **European standards groups:** Two standards-making bodies in the European Community are: the European Committee for Standardization (CEN) and the European Committee for Electrotechnical Standardization (CENELEC).

6. **Preservation microfilming:** For preservation filming see The Preservation Microfilming Handbook, published by the Research Libraries Group, Mountain View, California. Another good book on the subject edited by Nancy Gwinn and published by the American Library Association entitled Preservation Microfilming: A Guide for Librarians and Archivists.

7. **Computers and equipment:** Other standards dealing with computers and computer-peripheral equipment that are important in configuring imaging and preservation systems include: network standards (TCP/IP, NETbios, OSI/ISO, etc.), interface standards (SCSI, EDSI, etc.), display standards (VGA, XVGA, etc.), and operating system standards (DOS, Windows, OS/2, UNIX, etc.).

8. **Books:** There are two important books referenced in the September issue of Imaging Technology Report which are recommended reading for all practitioners on standards issues:

   "Document Imaging Standards Development: How, Why and For Whom?" (L034-1992)
   "Imaging Standards" (L001-1992)

Both are available from the AIIM bookstore.

NOTES

1 Nugent, William R., "Applications of Digital Optical Discs in Library Preservation and Reference," AFIPS -- Conference Proceedings, Vol. 52, pp. 771-775.

2 Nugent, William R., "Research in Extending the Longevity of Information on Digital Optical Disks and Videodiscs," Summaries: Electronic Imaging 86, Boston Mass., Nov. 86, pp. 790-795.

3 Ibid.

4 "National Archives Storage under Scrutiny," **Computerworld**, Sept. 1, 1986, page unknown.

5 Gilheany, Stephen J., "Requirements for an All Digital Engineering Data Management System," ATI Conference on Engineering Data Management, November 1983, pp. 2-7.

6 **RLG Preservation Microfilming Handbook**. Nancy E. Elkington, editor, March 1992. The Research Libraries Group, Inc.,

7 Bourke, Thomas, "Research Libraries Reassess Document Preservation Technologies," **Inform**, September 1990, pp. 30-34.

8 Frank, John W., "Micrographics and Optical Disc--Friends or Foes?," **IMC Journal**, July/August 1988, pp. 7-9.

9 Magnell, Glenn, "Micrographics and Optical Disk Technology: A Synergism in Information Management," **Image Update**, Issue 11, June 1989, pp. 1-4.

10 Magnell, Glenn, Michigan Chapter AIIM meeting held on March 28, 1989.

11 Black, David, "The New Breed of Mixed-Media Image Management Systems," **IMC Journal**, January/February 1989, pp. 9-13.

12 Moore, Frank A., "Spelling Out the Benefits of Imaging," **Inform**, February 1990, pp. 29-32.

13 Willis, Don, "The Future of CD-ROM," **IMC Journal**, Issue 2, March/April 1989, pp. 11-14.

14 Waters, Donald J., "From Microfilm to Digital Imagery," a report published by the Commission on Preservation and Access, June 1991.

15 Andrews, Christopher, "Mastering The CD-ROM Mastering And Replication Process", **CD-ROM Professional**, July 1991, pp. 17-18.

16 Lesk, Michael, "Image Formats for Preservation and Access," a report published by the Commission on Preservation and Access, July 1990.

17 Datapro Research Group, "Datapro Reports on Document Imaging Systems," **Document Imaging Systems**, February 1991, Vol. 2, No. 2, Section 5, pp. 3-4.

18 Datapro Research Group, "Datapro Reports on Document Imaging Systems," **Storage Technology & Products**, June 1992, Vol. 3, No. 6, Section 8, pp. 12-27.

19 Ibid.

20 Hawken, William R., "Copying Methods Manual, **Library Technology Program,** American Library Association, 1966, p. 30.

21 Gordon, Max, "Dots and Spots: Taking Care of EP&P Halftone Requirements," **Electronic Publishing & Printing**, November 1989, pp. 33-40.

22 Gilheany, Stephen, op. cit., pp. 10-11. (This paper and "Specifying a Digital Engineering Document Management System," Nuclear Records Management Association Annual Symposium, September 1984, also by Stephen Gilheany, are valuable sources of information on both film and digital image document storage and retrieval systems and imaging questions in general.)

23 Gilheany, Stephen J., op. cit.

24 Gordon, Max, op. cit.

25 Ibid.

26 Smith, Ross, "'G' Controller for Graphics Grayscaling," **Publishing**, March 1989.

27 Tapper, G.D., "Optical Discs--Standards," **IMC Journal,** July/August 1988, pp. 41-42.

28 Courtot, Marilyn, "Opening the Berlin Walls," **Inform**, March 1990, pp. 28-33.

FIGURES

Attributes of Micrographics

Advantages

- relatively low cost
- recognized archival medium
- inexpensive reader
- most cost-effective grayscale storage
- accepted as a legal medium
- excellent compaction
- standards for creating, processing, duplicating, storing, and reading exist

Disadvantages

- slow retrieval speed
- use can cause wear
- integrity of manual files is a problem
- single-user access
- less than ideal output quality
- resolution loss with succeeding copies

Figure 1

Attributes of Digital Imaging

Advantages

- excellent record access, distribution, and transmission
- multi-user simultaneous access
- file integrity
- improved quality possible through electronic image processing (restoration and enhancement)
- high-quality printed output
- no degradation on successive copies (each copy is as good as the original copy)
- easily reformatted (cut and paste)
- OCR to text possible
- electronic links to provide retrieval of individual pages

Disadvantages

- relatively high but decreasing cost
- relatively new technology
- permanent, but not archival, storage medium
- not yet accepted as legal reproduction
- implementation and operating costs increase in direct proportion to quality of captured image (resolution)

Figure 2

Attributes of Optical Disc

Advantages

- high speed retrieval
- longevity>20 years
- preserves file integrity
- excellent compaction
- multi-user access
- no wear during usage (non-contact read)
- excellent prospects for permanent (not archival) storage

Disadvantages

- high (but declining) cost
- relatively expensive retrieval systems required
- not yet cost-effective for storage of grayscale page images
- not yet accepted as legal document storage medium
- new or no standards

Figure 3

Resolution Test Targets

(one of several types)

resolution test patterns

Figure 4

Effective Use of Images Storage

Hierarchical Storage Concept

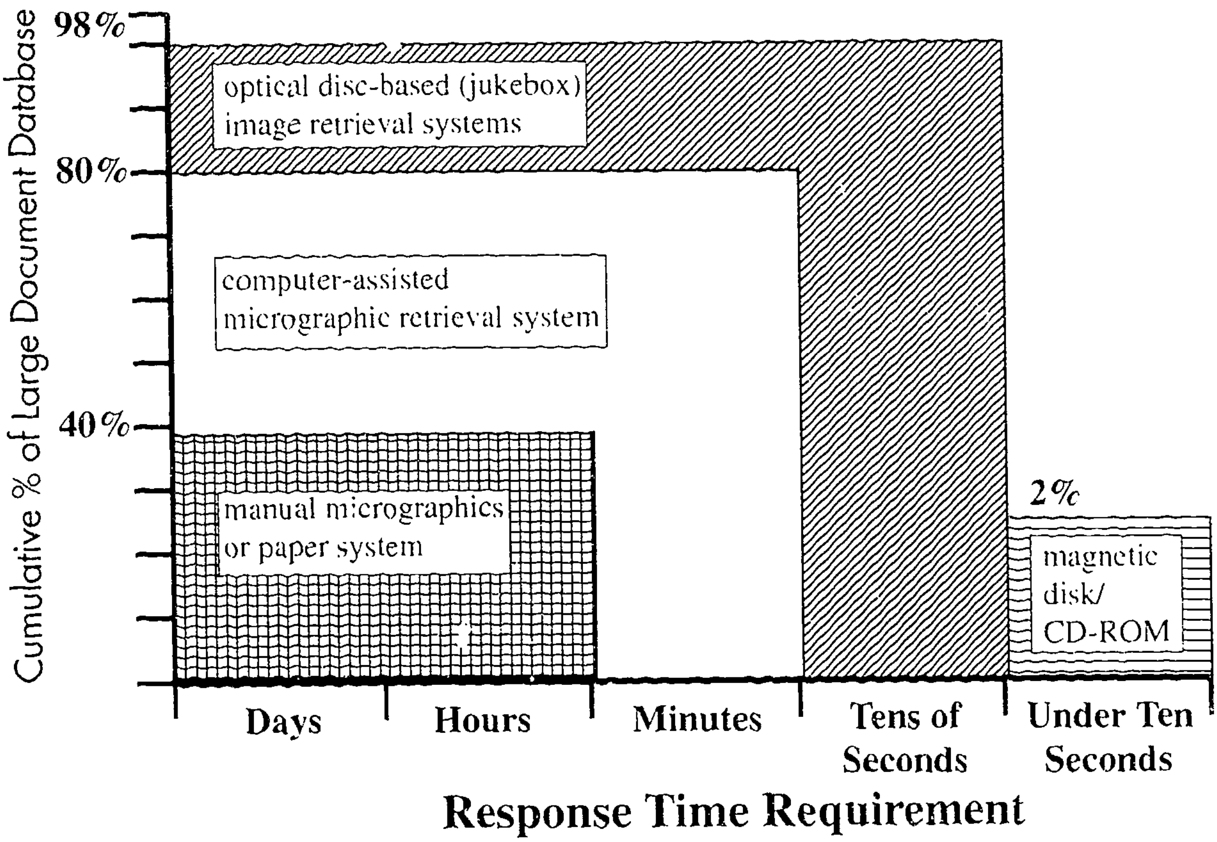
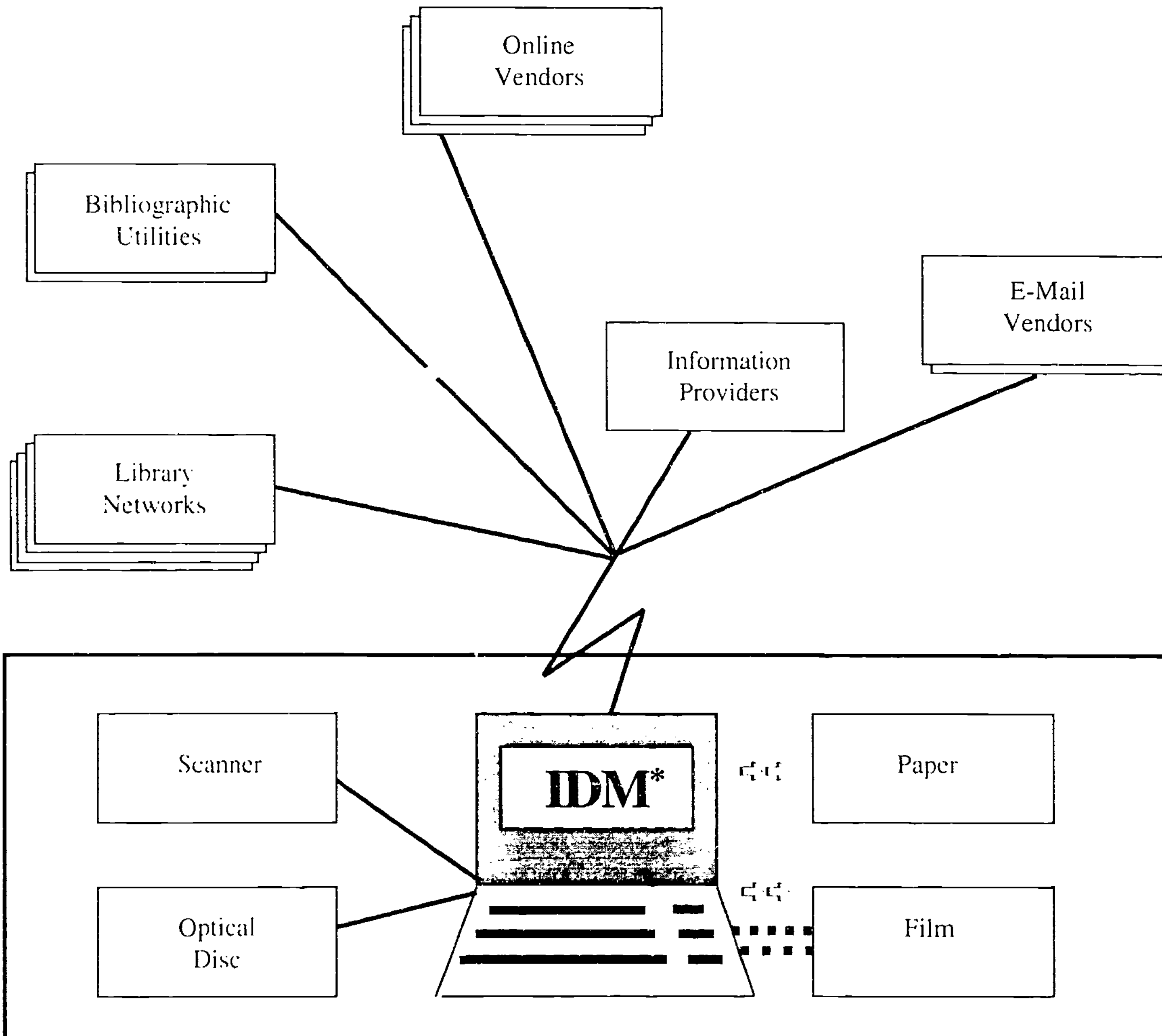


Figure 5

The Hybrid End-User Access System

Single Point of Access



*Information Delivery Machine

Figure 6

Information Storage Requirements

Various Formats

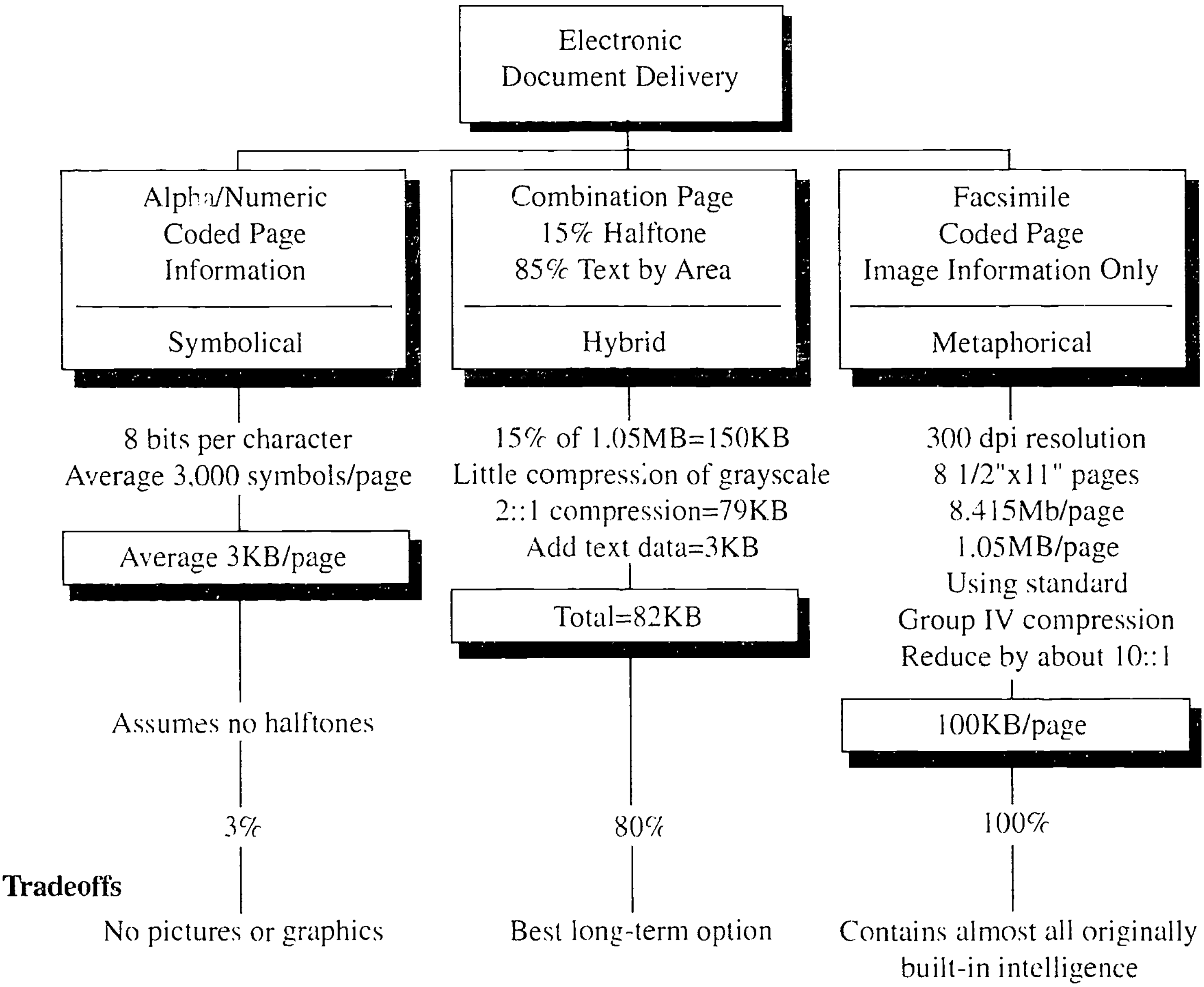


Figure 7

Digital Image System

Components

- Database Server—could be the same as the hybrid end-user access system
  - —database software
  - —temporary magnetic storage
  - —permanent storage (optical, 8mm videotape, 4mm DAT, etc.) (optional)
  - —network interface boards, cable, and software (optional)
  - —compression/decompr ssion hardware, software

- Scanner(s)

- Workstation(s)
  - —application software
  - —compression/decompression hardware or software
  - —local (temporary storage)
  - —high-resolution display (optional)

Figure 8

Preservation Microfilming System

Single Camera, Low-Speed Processor—Costs to Purchase

Item	Cost	#	Total Cost
Camera (used)	$20,000	1	$20,000
Camera (new)	$100,000		
Book cradle	$5,000	1	$5,000
Film processor(slow)	$15,000	1	$15,000
Film processor (med)	$40,000		
Film processor (high)	$100,000		
Densitometer	$3,000	1	$3,000
Microscope	$1,500	1	$1,500
Film printer	$10,000	1	$10,000
Inspection reader	$2,000	1	$2,000
Misc. inspection equipment	$400	1	$400
Film winders	$500	1	$500
Ultrasonic splicer	$3,000	1	$3,000
Darkroom support equipment	$3,000	1	$3,000
Sensitometer	$2,000	1	$2,000
Darkroom supplies	$550	1	$550
Plumbing	$1,000	1	$1,000
A/C & humidity	$3,00	1	$3,000
Construction & supplies	$4,000	1	$4,000
	Total		**$73,950**

Figure 9a

Multiple Camera, Low-Speed Processor—Costs to Purchase

Item	Cost	#	Total Cost
Camera (used)	$20,000	5	$100,000
Camera (new)	$100,000		
Book cradle	$5,000	5	$25,000
Film processor (slow)	$15,000	2	$30,000
Film processor (med)	$40,000		
Film processor (high)	$100,000		
Densitometer	$3,000	2	$6,000
Microscope	$1,500	1	$1,500
Film printer	$10,000	2	$20,000
Inspection reader	$2,000	2	$4,000
Misc. inspection equipment	$400	2	$800
Film winders	$500	2	$1,000
Ultrasonic splicer	$3,00	2	$6,000
Darkroom support equipment	$3,000	1	$3,000
Sensitometer	$2,000	1	$2,000
Darkroom supplies	$550	2	$1,100
Plumbing	$1,000	2	$2,000
A/C & humidity	$3,000	2	$6,000
Construction & supplies	$4,000	2	$8,000
	Total		**$216,400**

Figure 9b

Preservation Microfilming System

Multiple Camera, Medium-Speed Processor—Costs to Purchase

Item	Cost	#	Total Cost
Camera (used)	$20,000	10	$200,000
Camera (new)	$100,000		
Book cradle	$5,000	10	$50,000
Film processor(slow)	$15,000	1	$15,000
Film processor (med.)	$40,000	1	$40,000
Film processor (high)	$100,000		
Densitometer	$3,000	2	$6,000
Microscope	$1,500	1	$1,500
Film printer	$10,000	3	$30,000
Inspection reader	$2,000	2	$4,000
Misc. inspection equipment	$400	2	$800
Film winders	$500	3	$1,500
Ultrasonic splicer	$3,000	3	$9,000
Darkroom support equipment	$3,000	2	$6,000
Sensitometer	$2,000	1	$2,000
Darkroom supplies	$550	3	$1,650
Plumbing	$1,000	3	$3,000
A/C & humidity	$3,000	3	$9,000
Construction & supplies	$4,000	3	$12,000
	Total		**$391,450**

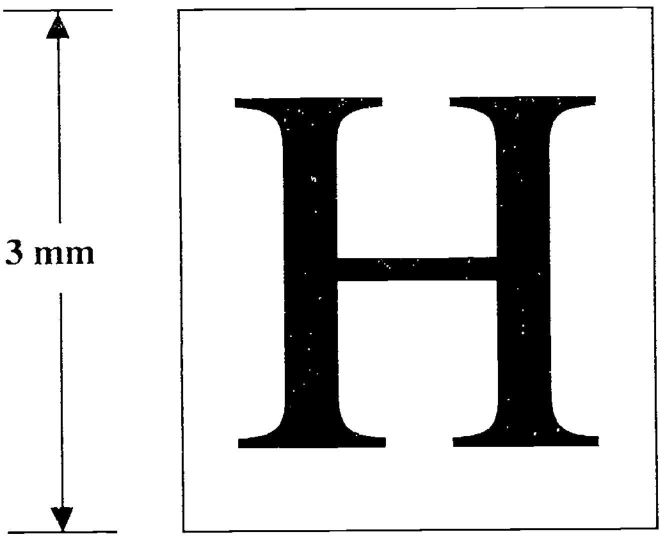
Figure 9c

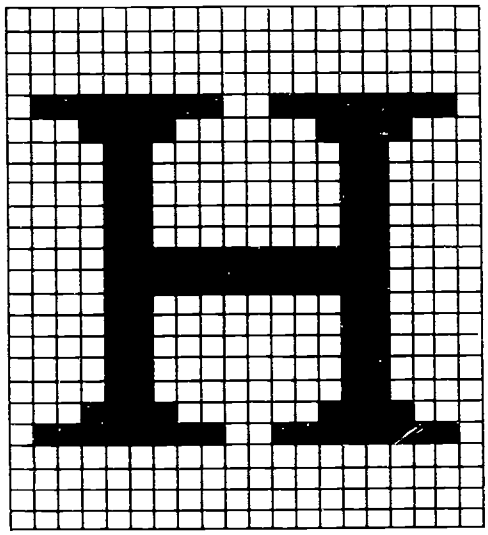
Multiple Camera, High-Speed Processor—Costs to Purchase

Item	Cost	#	Total Cost
Camera (used)	$20,000	20	$400,000
Camera (new)	$100,000		
Book cradle	$5,000	20	$100,000
Film processor (slow)	$15,000	1	$15,000
Film processor (med.)	$40,000		
Film processor (high)	$100,000	1	$100,000
Densitometer	$3,000	3	$9,000
Microscope	$1,500	1	$1,500
Film printer	$10,000	4	$40,000
Inspection reader	$2,000	2	$4,000
Misc. inspection equipment	$400	3	$1,200
Film winders	$500	4	$2,000
Ultrasonic splicer	$3,000	4	$12,000
Darkroom support equipment	$3,000	2	$6,000
Sensitometer	$2,000	1	$2,000
Darkroom supplies	$500	4	$2,200
Plumbing	$1,000	4	$4,000
A/C & humidity	$3,000	4	$12,000
Construction & supplies	$4,000	4	$16,000
	Total		**$726,900**

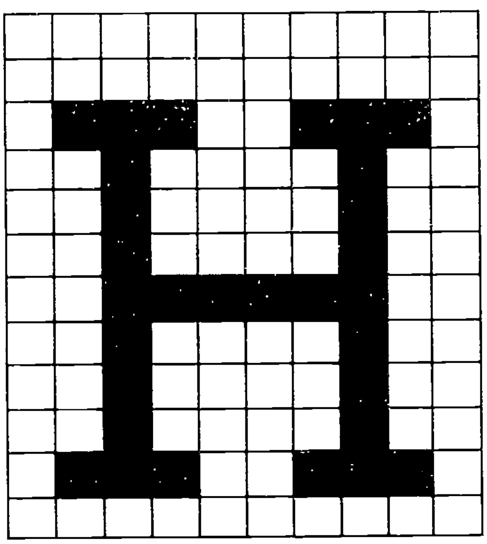
Figure 9d

Resolution

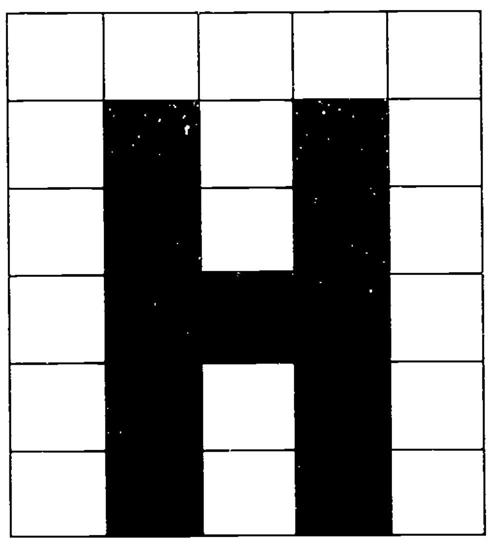




8 Lines per mm (200 dpi)



4 Lines per mm (100 dpi)

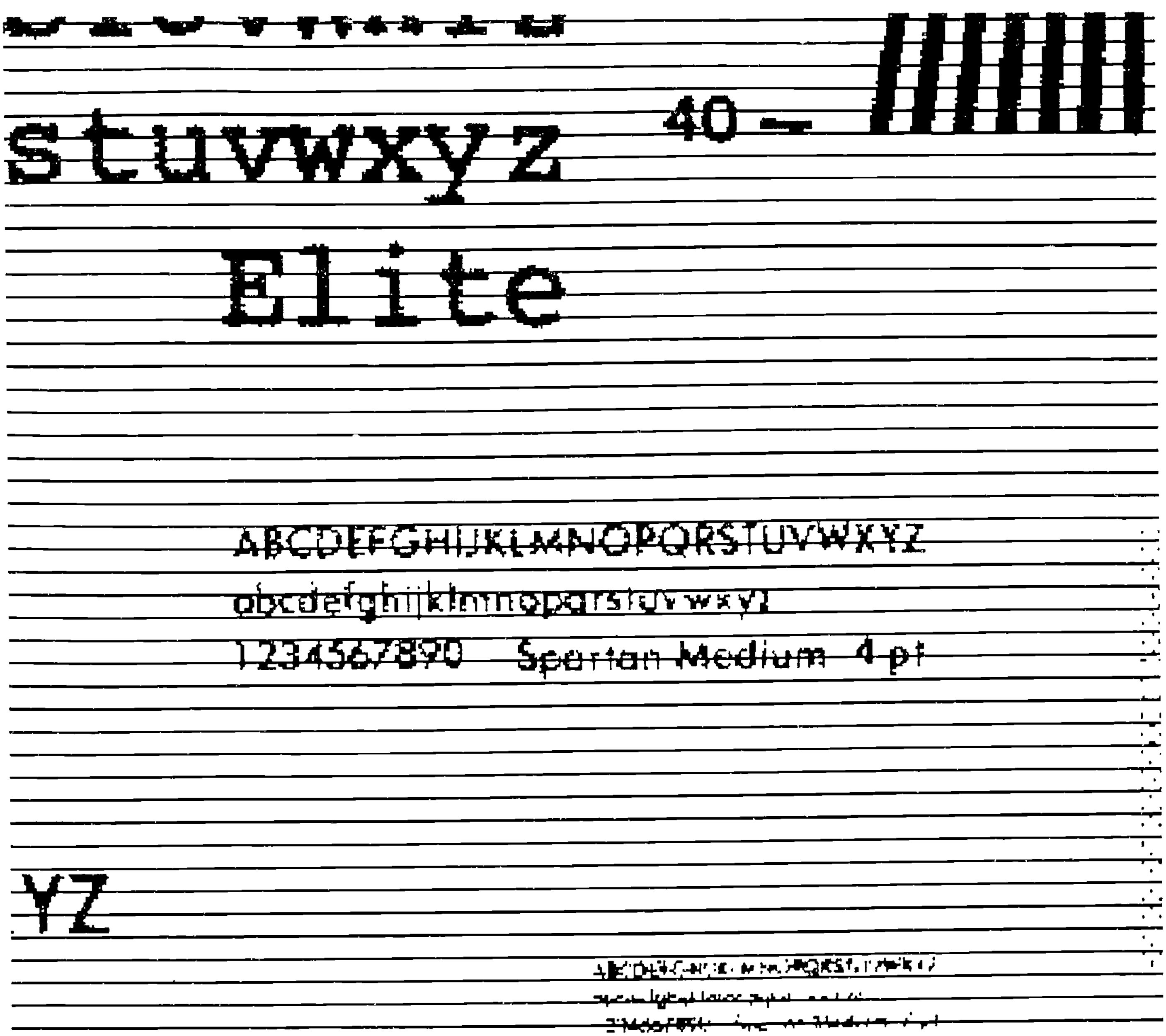


2 Lines per mm (50 dpi)

Figure 10

Portion of 300-dpi Page

(enlarged to show the interaction of scanning resolution on character size)



Taken from scan test chart

(See Figure 3)

Figure 11a

Measurements of Type Design

copying methods manual

Figure 11b

Portion of 400-dpi Page

(enlarged to show the interaction of increased scanning resolution on character size)

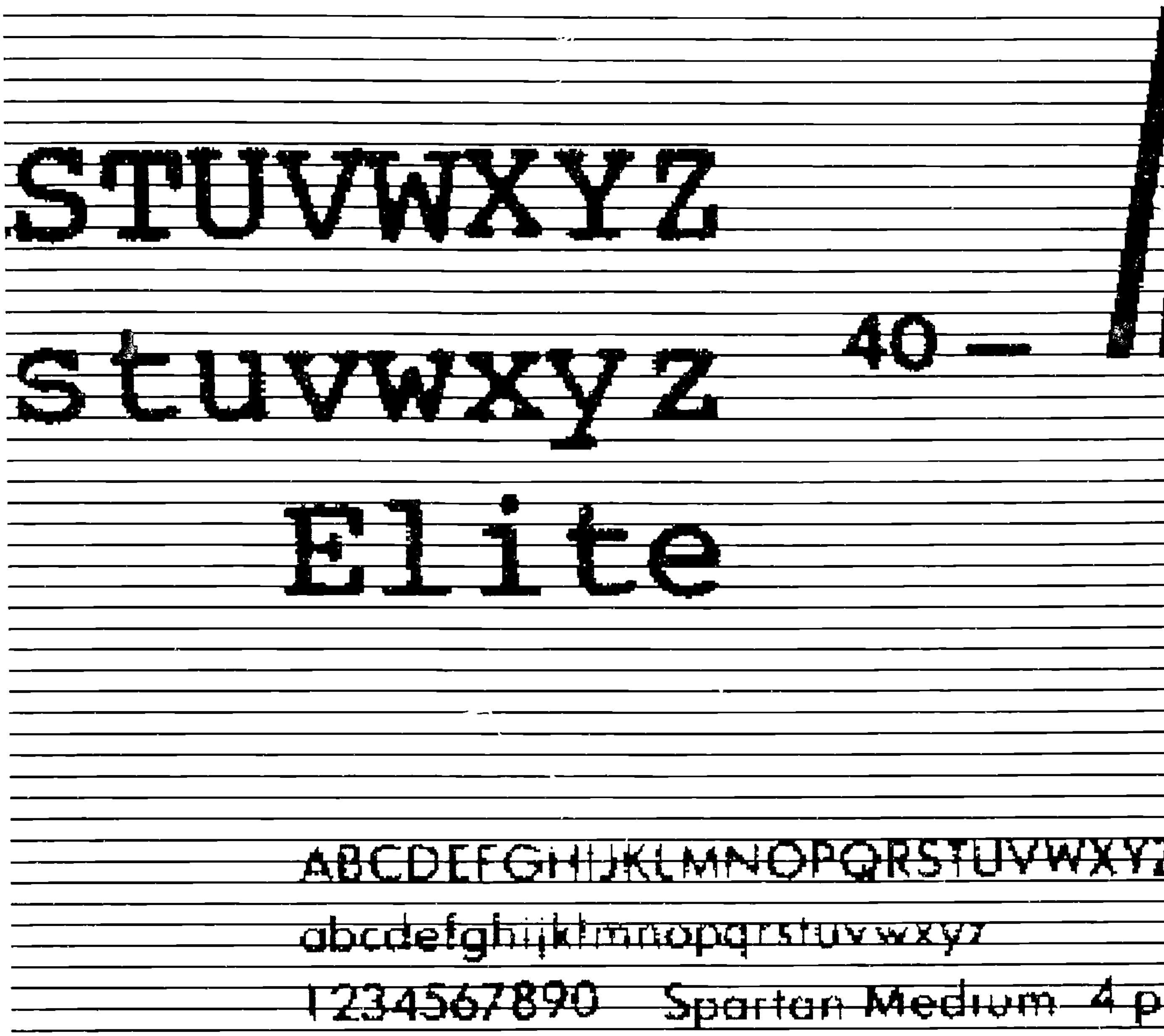


Figure 12

Grayscale vs. Resolution

(example of tradeoffs between screen ruling [resolution] and number of gray levels in a halftone image)

30 lines per inch diagonally,
7 x 7 dot combination,
rendering 101 gray levels

35 lines per inch diagonally,
6 x 6 dot combination,
rendering 74 gray levels

53 lines per inch diagonally,
4 x 4 dot combination,
rendering 33 gray levels

70 lines per inch diagonally,
3 x 3 dot combination,
rendering 19 gray levels

Figure 13

Image Enhancement Example

Before

binary image -
no image enhancement

After

binary image —
image enhanced page

300 dpi resolution

Figure 14

Image Enhancement Quality Comparison

original image

image made from
office copier

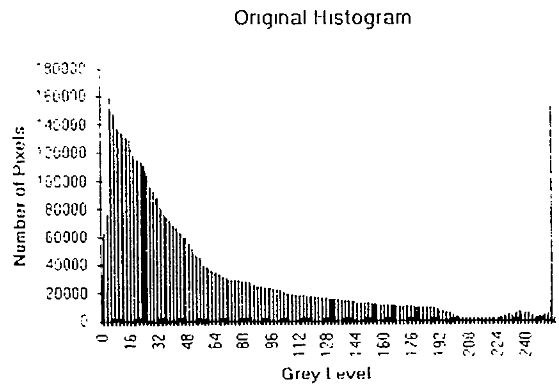
image made by
scanning in binary mode
300 dpi

enhanced image
300 dpi

Figure 15a

Example of Rehistogramming in Image Enhancement

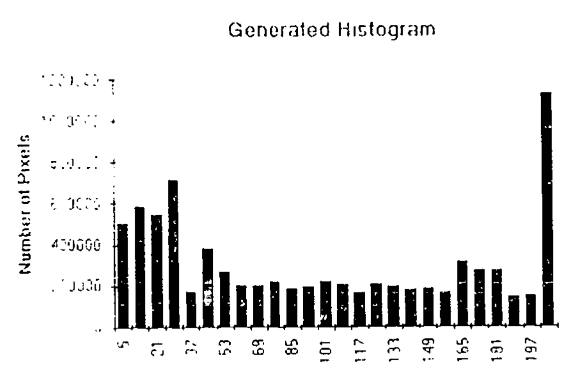
scanned image after
image enhancement—300 dpi



original histogram
of grayscale values

original

after image enhancement and
rehistogramming—300 dpi



reallocated grayscale values

Figure 15b

Comparison: Film vs Digital Resolution

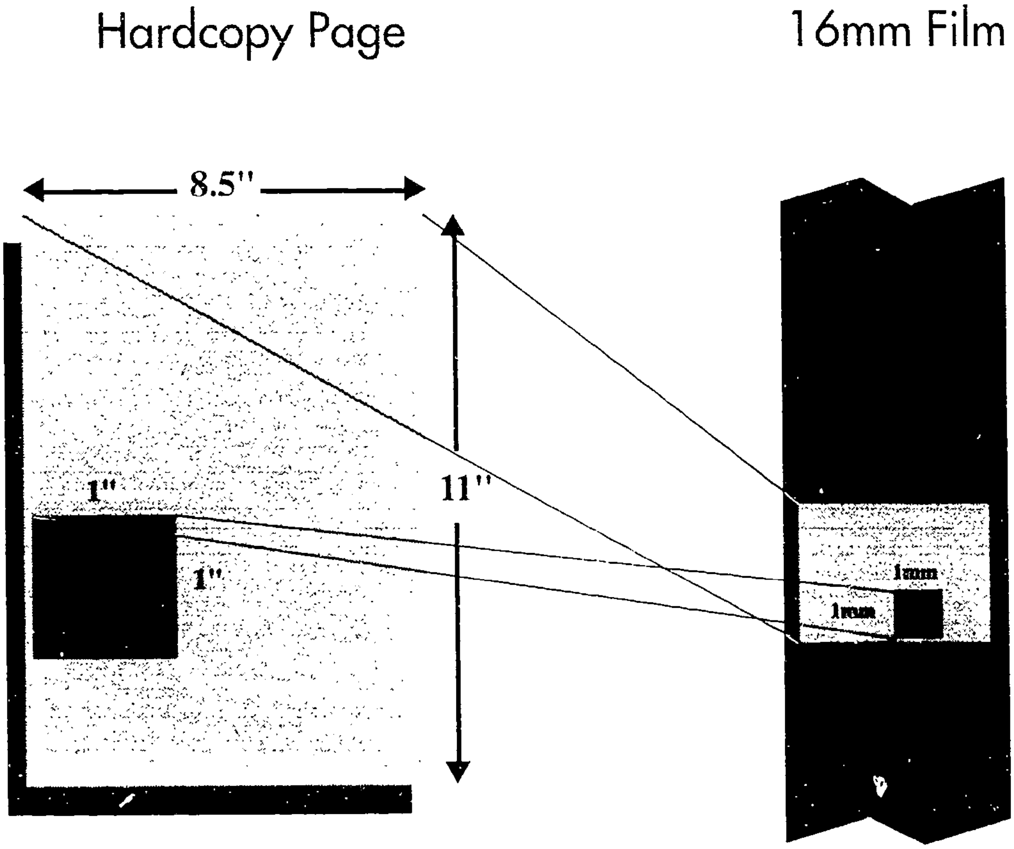
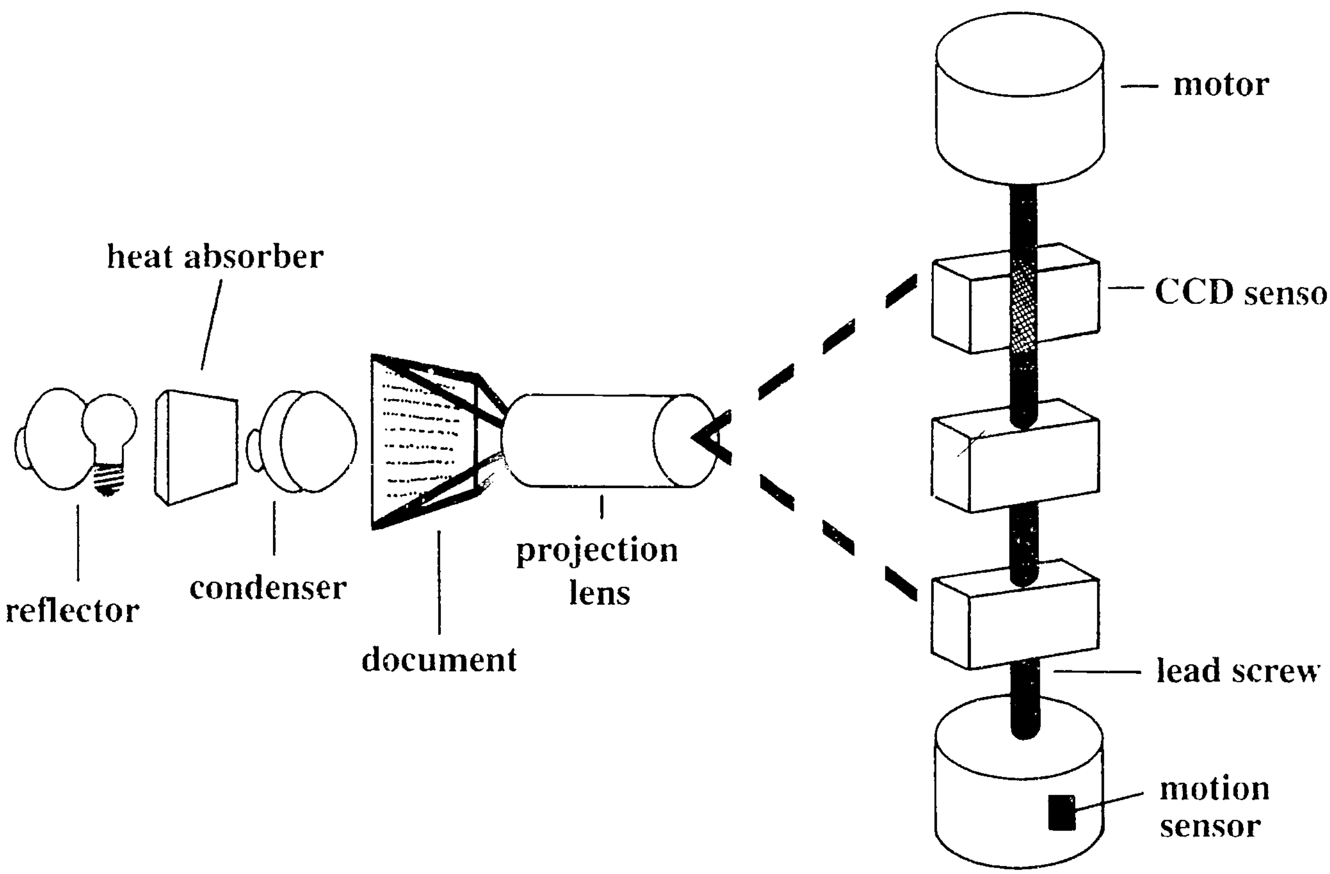


Image on film 25x smaller
Reduction ratio = 25x
but
1 inch = 25.4 millimeters

Figure 16

Diagram of Charge Coupled Device Scanner



used with permission of G. Walters,
Rothchild Consulting

Figure 17

Printer Halftone Example

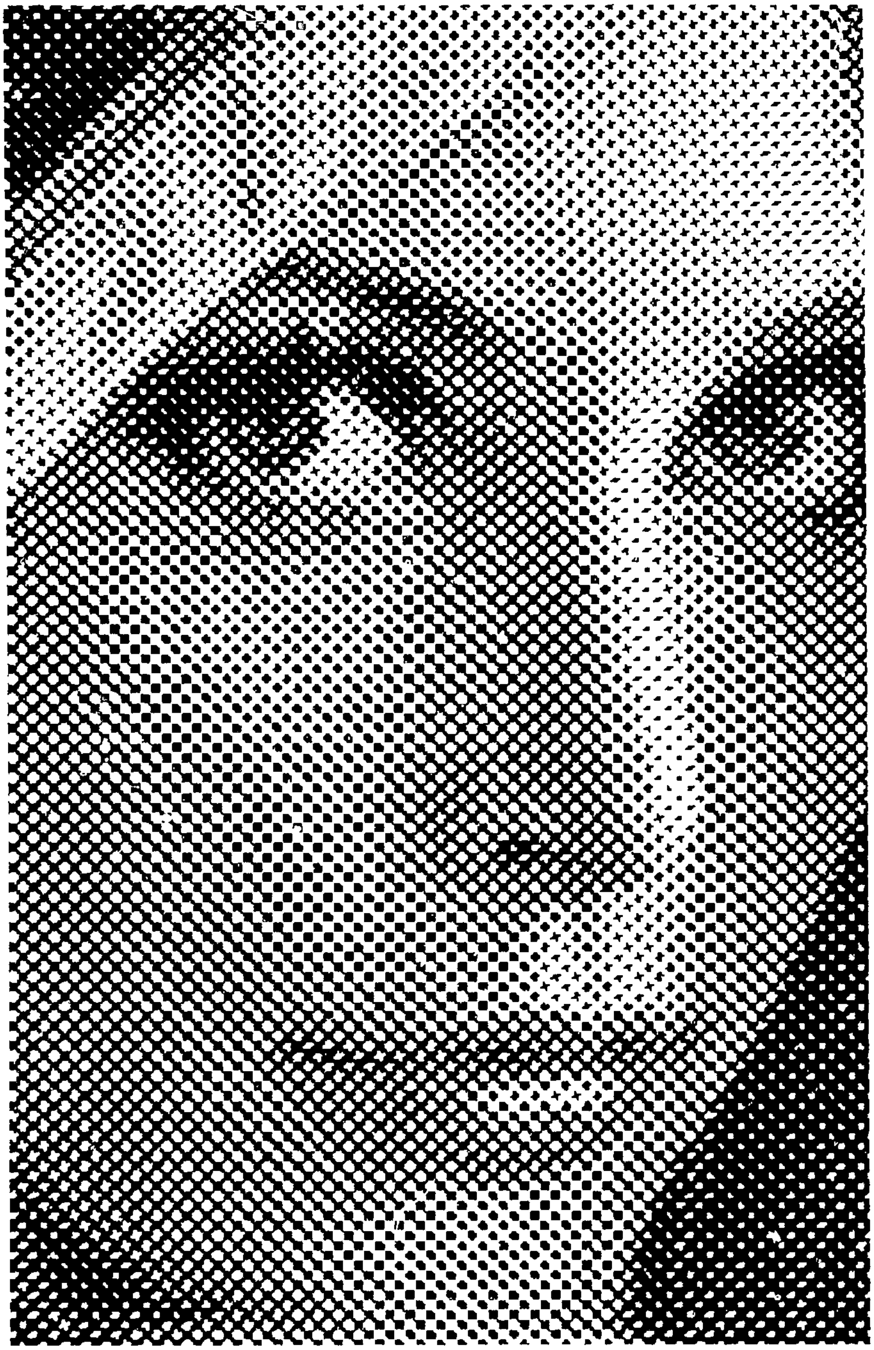


illustration of halftone cells
created on a standard
300-dpi printer

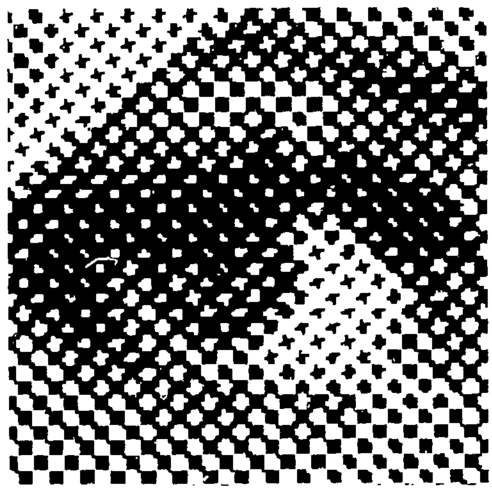


Figure 18

Example of Effects of Halftoning on Printed Images

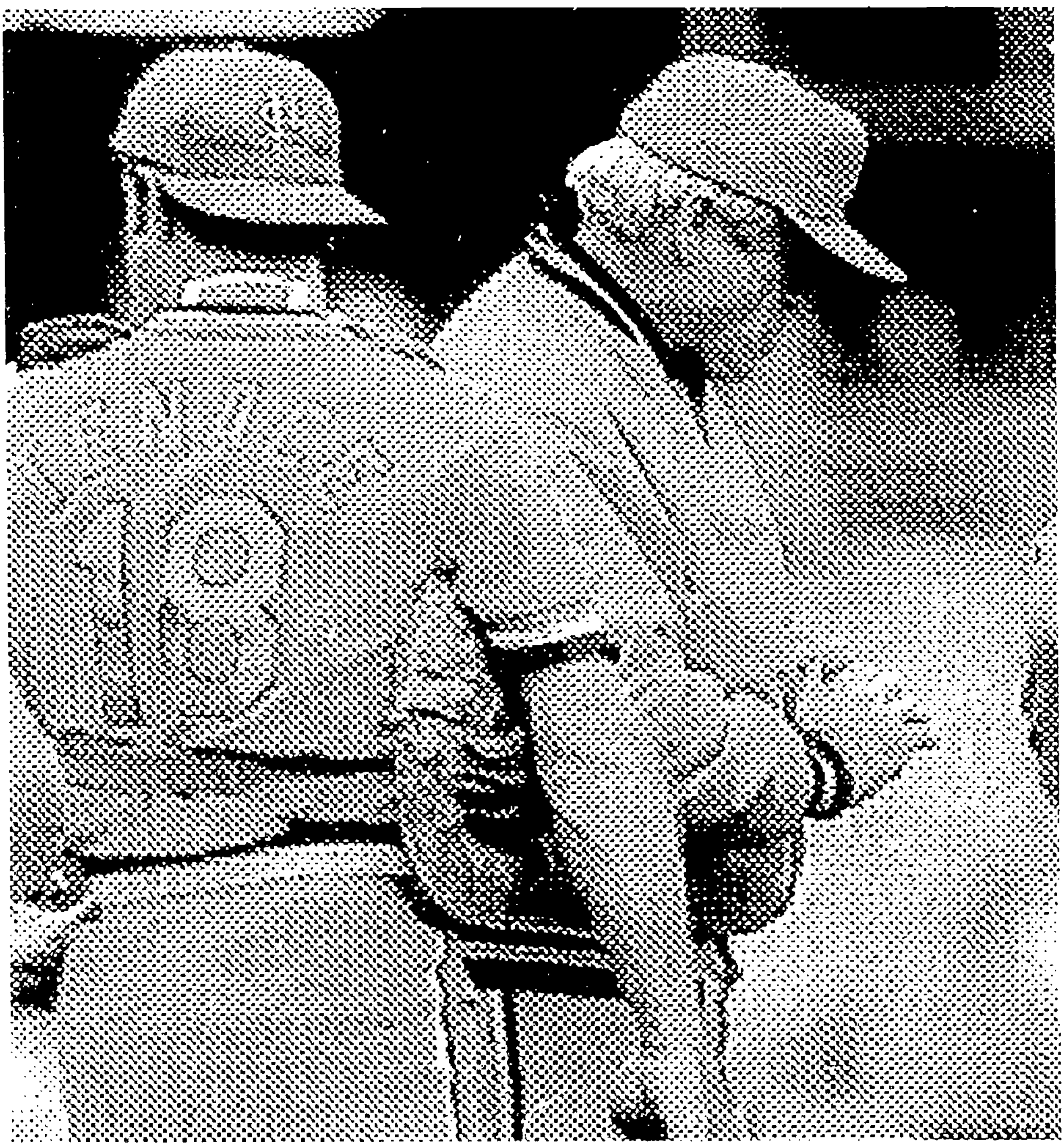


Figure 19