Working with Microfilm

Paul Conway Yale University Library

Preservation microfilm can be an excellent source-medium for digital conversion projects if certain caveats are taken into consideration. This section describes what librarians and archivists need to know about working with existing microfilm to produce high-quality digital images that can be displayed as images and/or processed with OCR conversion software.

Background -- Project Open Book

Microfilm has been used as a medium for preservation and access since the 1930s. By the middle of the 1980s, international standards fully defined the archival qualities of preservation microfilm (Fox, 1996). The Research Libraries Group, working in close association with the American Library Association, established procedures for creating film that meets or exceeds archival standards (Elkington, 1992). By the end of 1999, the National Endowment for the Humanities had provided partial support for the preservation of over 800,000 brittle volumes on microfilm. The nation's collection of preservation microfilm is the first and one of the largest *virtual libraries* in the world (Conway, Selecting,1996).

In the early 1990s, Don Willis, one of the most prominent experts on the creation of preservation microfilm, proposed that it was technologically and economically feasible to create high-quality digital images from microfilm (Willis, 1992). At the time he wrote, few people outside the commercial sector -- and no U.S. archivists or librarians -- were in a position to test the hypotheses that Willis proposed. The conversion of microfilm was largely confined to corporations that needed to convert *legacy files* from microfilm (typically, case files and standard office documents) on a highly selective basis. What was needed was a systematic exploration of the issues associated with tapping the content of hundreds of thousands of brittle books, newspapers, and serials preserved on 35 millimeter microfilm now housed in research libraries and archives around the country. If it proved feasible to obtain high quality images at a reasonable cost from the nation's corpus of preservation film, then this material could be added to what was then expected to be a national digital image resource.

Yale University Library, with the assistance of the Commission on Preservation and Access and the National Endowment for the Humanities, accepted the job of developing a sequence of projects, collectively termed Project Open Book, to test Don Willis's hypotheses (Waters, 1991). Yale designed and implemented Project Open Book in close association with Cornell University, which at the time was also deeply engaged in digital imaging R&D, using books as the principal conversion source. Yale adopted Cornell's recommendations for base line image quality and then went on to develop a complex cost study to test the underlying economic assumptions of the imaging process. Project Open Book defined the relationship between quality and cost. The project established rules of thumb for maximizing quality and baseline cost estimates for the microfilm conversion process (Conway, "Yale," 1996).

Since the Yale project has been completed, additional projects have contributed to the general microfilm-scanning knowledge base. Additionally, several service bureaus have

begun offering conversion services to libraries and archives. These commercial organizations are able to meet or exceed quality expectations at a costper-image that is not as low as the benchmarks identified by Yale, yet still fairly cost effective. In 1999, the principal investigators of the Cornell and Yale projects
pooled their knowledge of the hybrid approach and developed a set of recommendations for converting brittle books from either film or the original item (Chapman, Conway & Kenney, 1999). Together, these developments make it possible to recommend best practices for certain kinds of materials on film and to identify when microfilm is not the best source.

Image Quality Considerations

Image quality is the first concern. High contrast 35-mm microfilm, produced according to ANSI/AIIM specifications to a Quality Index level of at least 5 (on a scale of 1 to 8) has the equivalent digital resolution of at least 800 dots per inch (dpi). It is not yet possible (nor may it be necessary) to achieve this level of scanning across the full width of the 35 mm microfilm frame. High resolution scanning from microfilm varies from 300 to 600 dpi. Bit depth ranges from bitonal (1 bit per pixel) to full gray (8 bits per pixel). Scanners for color roll film (a relative rarity in libraries) are not available commercially, although such technology is an important part of the movie industry (Kenney & Chapman, 1996).

Because of the high risk of damage, master microfilm negatives (1N) should never be used as a scanning source. Research at Yale and in Germany has shown that the same level of image quality can be obtained from a duplicate negative (2N) without placing the master negative in jeopardy (Weber, 1997). If only a positive use copy (3P) is available, it is possible to obtain a readable digital image, although some detectable drop-off in image quality should be expected.

Characteristics of the original source document and characteristics of the microfilm medium strongly influence the quality of the individual images and the total image product. Here are some highlights.

Characteristics of the Original Source (e.g., book, document, print, map)

- High contrast between text (ink) and surface (paper) yields best results
- Discolored, damaged, uneven edges of paper complicate scanner setup
- Bleed-through of text from verso limits threshold options
- Foxing, mold, stains, and fire and water damage may be accentuated by scanning
- Tight gutters in bound volumes distort film and digital imagery unless corrected
- Fold-outs and oversize inserts may not be represented in digital form as accurately as baseline document (in-line modifications to scanner setting are require

Characteristics of the Microfilm

Image Quality

- Polarity: negative microfilm yields higher quality images than positive film
- Density: medium contrast (dMax ca. .90) to high contrast (dMax ca. 1.30) film results in higher quality images than low contrast (dMax ca. .80) negatives. RLG dMin guideline (< .25) holds.

- Reduction ratio: lower is better; accurate recording of ratio is crucial for reproduction at original size
- Skew: minimize or eliminate -- no greater than 5 degrees from parallel

Product Quality

- Consistent placement: minimize or eliminate *centerline weaving*
- Duplicate images: duplicate images bracketing illustrations have minimal impact
- Splices: eliminate splices inside a given volume on the reel
- Dimensions of original: record accurately on bibliographic target
- Blank frames: eliminate or reduce quantity wherever possible
- Orientation: A2 position provides most consistent product with some scanners; one full frame per image is generally preferable.
- Test charts: incorporate RIT Alphanumeric Test Chart into scanner setup routine

Bottom Line on the Quality of Bitonal Scanning

Conversion Cost Issues

Imaging costs are driven by scanner pricing structures, labor costs, and the overall speed of the conversion system. The throughput speed of a given scanner is a product of at least three factors:

- Image resolution (the lower the resolution the faster the output)
- Electrical engineering (fast refresh rate of the CCD array and fast data transfer rate equals fast output)
- Mechanical engineering (more rigorous film transport mechanisms provide for quicker throughput). It is somewhat difficult to compare scanner speeds by studying manufacturer specifications. In its complex study, Project Open Book examined the cost of the imaging process in terms of equipment and human resources (Conway, *D-Lib Magazine,* 1996). The cost model factored in the actual costs of hardware, software, integration support, and optical storage media and then converted these costs to a range of per-book and per-image costs. Most importantly, the Yale study assessed costs for each of the processing steps of the conversion process.

The Yale study identified a number of factors that contribute to variation in costs, including the following:

- The impact of original source and microfilm characteristics varies among process steps.
- Most time-consuming conversion steps (scanning in continuous mode, indexing, scanner setup, and file transfer) are not greatly influenced by original source or microfilm factors.
- Original source characteristics influence costs more than microfilm characteristics.
- Original source and microfilm characteristics, combined, have dramatic impact on quality but only marginal impact on costs.
- Pre-scan inspection of microfilm (a relatively inexpensive processing step) is an important mechanism for predicting quality control challenges, but is not sufficient for identifying significant scanning and indexing complexities that
- Nature, quality, and value of complex illustrations determine the appropriateness of bitonal scanning; if illustrations are vital and complex, then bitonal scanning may not be appropriate.
- Crispness of text (printed or hand-written) is essential for legibility of the digital image.

- No appreciable improvements occur in image quality with continuous tone film scanned in bitonal mode arise during the conversion process.

Service Bureaus

Vendors can do the hard work. It is not necessary to purchase microfilm scanning hardware and software for in-house use in order to accomplish the conversion of microfilm. A number of companies in the United States offer conversion services,

Characteristics of the Original Source (e.g., book, document, print, map)

- Characteristics of the original source that have a large impact on quality (*e.g.,* faded text, bleed through) have little impact on the cost of digital conversion.
- The number of pages in the chunk of material being scanned has a significant financial impact on all conversion processes.
- Books without tables of contents or page numbers pose significant indexing challenges (and thus higher costs), but also complicate prescan inspection and all aspects of quality control.
- The presence of illustrations is only one of many factors that combine to explain variation in the cost of the most time-consuming processing steps.
- The costs of quality control processes carried out during scanning, indexing, and final acceptance are strongly influenced by original source characteristics (*e.g.,* tight gutter margins, cropped text, illustrations).
- Preparation of a bound volume prior to microfilming (*e.g.,* disbinding, careful cropping) can significantly reduce the cost of setting scanner parameters.

Characteristics of the Microfilm

- Reduction ratio is the single most important microfilm characteristic influencing costs. The smaller the ratio the lower the conversion cost.
- Skewed microfilm images, an all-too-common factor, increase the cost of scanning, quality control, and inspection.
- Splices inside a given volume influence the cost of several important steps, but occur too infrequently to matter much.
- The cost-per-item of scanner set up is not influenced by any characteristics of microfilm.
- Density variation has no impact on the cost of conversion.
- Investment in better quality microfilm has only marginal cost-reduction benefits.
including:
- Preservation Resources of Bethlehem, PA <http://www.oclc.org/oclc/presres/index.htm>
- Northern Micrographics of La Crosse, WI <http://normicro.com>, and
- microMedia Imaging Systems, Inc. of Lake Success, NY.
- sources for information on service bureaus are:
- *Imaging Magazine* <http://www.imagingmagazine.com> and the Association for Information and Image Management (AIIM) <http://www.aiim.org>.

You must be a member ($125 individual) to take advantage of AIIM's excellent library and referral services. It is very important to test the products (*deliverables*) of a service bureau before finalizing a contract. Most service bureaus will conduct scanning tests for free or for a modest fee as part of a competitive bidding process. It is your responsibility to specify the

quality level of the digital images in terms of resolution, dynamic range (bit depth), and postscan image processing (*e.g.,* deskew, despeckle, and tone adjustment). It is also your responsibility to specify whether it is acceptable for the vendor to use equipment that uses synthetic resolution tools to offset the resolution limitations of the equipment. Finally, it is also your responsibility to specify the characteristics of the output files in terms of file format, naming conventions and directory structures, and delivery mechanism (*e.g.,* CD-ROM, FTP server, magnetic tape).

Equipment Options

If you are working with a contractor to accomplish your imaging goals, it will not be necessary to purchase scanning equipment. Nevertheless, you can and should learn as much as you can about the capabilities of scanning equipment by contacting the manufacturers of hardware and software systems.

Hardware/software capabilities must be understood in order to develop quality and cost specifications, regardless of whether a scanning program is carried out in the library. Scanning results will vary across machines, however, depending on how the software for a given machine defines the thresholds (analogous to contrast settings on a photocopier), sets the various filter options, and applies various algorithms that interpret and adjust pixel encoding. The more that is known about how the scanner interprets and codes what it sees, the better the resulting images.

Ultimately, quality specifications, technology capabilities, and the visual characteristics of the original source combine to determine the quality and cost of the image product.

The following five companies either manufacture or resell four systems for microfilm scanning in the United States. In general, hardware and software are bundled as a single package. The amount of customization that can be specified by the buyer for either hardware or software varies from none (Minolta) to extensive (Amitech). The amount of end-user control over the equipment also varies widely. It is important to view and test equipment in real-world settings before purchasing equipment. The best way to undertake this testing is to ask hardware companies for a short list of client-references in the area and then contact these references directly.

Amitech Corporation <http://www.amitech.com>
5501 Backlick Road
Suite 200
Springfield, VA 22151
Phone: 703-256-2020 Fax: 703-256-9153
Amitech resells three of the four microfilm scanners (Mekel, SunRise, Wicks & Wilson) that are presently available and also provides a variety of software packages (customizable) that control the scanner operation and carry out various postscan data management tasks (*e.g.,* deskew, despeckle, compression).

Mekel Engineering, Inc. <http://www.mekel.com>
2800 Saturn Street, Suite B
Brea, CA 92821-6201
Phone: 714-996-5600 Fax: 714-996-5696

The Mekel M500 is the premier high-speed microfilm conversion product. It is capable of handling 35 mm or 16-mm roll film. The Mekel M560 is the associated hardware for fiche scanning.

Minolta Corporation <www.minolta.com>
101 Williams Drive
Ramsey, NJ 07446
Phone: 800-964-6658 Minolta manufactures the MS 3000 Microform Scanner, which can handle a full suite of formats if the transport mechanism is changed. The scanner is highly automated and provides limited operator flexibility.

SunRise Imaging, Inc. <http://www.sunriseimg.com>
1250 N. Tustin
Anaheim, CA 92807
Phone: 714-632-2160 Fax: 714-632-2161

The SunRise ProScan III is the most complex and comprehensive microfilm scanner on the market. It converts in both bitonal and gray scale mode and can handle a variety of formats depending on the configuration of the film support mechanisms.

Wicks & Wilson, Inc. <http://www.amitech.com>
Morse Road Basingstoke
Hampshire RG226PQ England
Phone: 011441256842211

The Wicks & Wilson 4000 and 4001 Scanstations are the newest arrivals to the U.S. market. They are manufactured in England by a company that specializes in high-tech imaging applications, such as virtual reality gloves. At publication time, the WW machines are available only through Amitech. The manufacturer claims high-resolution scanning and ease of use are key features.

Further Research Needed

Research needs to be done to certify the feasibility of converting nonbook materials, especially newspapers and manuscripts. Additionally, the challenges of working with microfilm that has not been created with rigorous archival standards are not well understood, including:

- Older film
- Scratched or damaged film
- 16 mm film
- Continuous tone film
- Positive polarity film
- Third generation film.

Conclusion

In the past decade, microfilm-scanning technology has matured to the point where you have distinct options regarding hardware and software capabilities, as well as choices about the quality of the end products and the cost of the technology. Quality is increasing; per-image costs are declining. You should have confidence that the digital image conversion of primarily text-based materials from preservation microfilm is both technically feasible and economically competitive with other
conversion technologies.

Sources

Chapman, Stephen, Paul Conway, and Anne R. Kenney. *Digital Imaging and Preservation Microfilm: The Future of the Hybrid Approach for the Preservation of Brittle Books.* Washington, DC: Council on Library and Information Resources, 1999. [Online] Available: http://www.clir.org/programs/cpa/hybridintro.html#description

Conway, Paul. "Selecting Microfilm for Digital Preservation: A Case Study from Project Open Book." *Library Resources & Technical Services* 40 (January 1996): 67-77.

-----. "Yale University Library's Project Open Book: Preliminary Research Findings," *D-Lib Magazine* (February 1996) [Online]. Available: http://www.dlib.org/magazine.html

------. *Conversion of Microfilm to Digital Imagery: A Demonstration Project.* New Haven, CT: Yale University Library, 1996.

Conway, Paul and Shari Weaver. *The Setup Phase of Project Open Book.* Washington, DC: Commission on Preservation and Access, June 1994. [Online]. Available: http://www.clir.org/pubs/reports/conway/conway.html.

Elkington, Nancy, ed. *RLG Preservation Microfilming Handbook.* Mountain View, CA: The Research Libraries Group, Inc., 1992.

Fox, Lisa, ed. *Preservation Microfilming: A Guide for Librarians & Archivists,* 2nd ed. Chicago: American Library Association, 1996.

Kenney, Anne R. and Stephen Chapman. *Digital Imaging for Libraries and Archives.* Ithaca, NY: Cornell University Library, 1996.

Watters, Donald J. *From Microfilm to Digital Imagery.* Washington, DC: Commission on Preservation and Access, June 1991.

Waters, Donald J. and Shari Weaver. *The Organizational Phase of Project Open Book.* Washington, DC: Commission on Preservation and Access, September 1992. http://www.clir.org/pubs/reports/openbook/openbook.html/openbook.html

Weber, Hartmut and Marianne Dorr. *Digitization as a Method of Preservation? Final Report of a Working Group of the Deutsche Forschungsgemeinschaft.* Washington, DC and Amsterdam: Commission on Preservation and Access and European Commission on Preservation and Access, 1997.

Willis, Don. *A Hybrid Systems Approach to Preservation of Printed Materials.* Washington, DC: Commission on Preservation and Access, 1992. http://www.clir.org/pubs/reports/willis/index.html