Human-Readable Preservation of Digital Images to Microfilm

Don Williams and Peter D. Burns
Eastman Kodak Company
Rochester, NY USA

Abstract

The predominant way of preserving digital images is in their current form. Doing so for the long term, however, will require format migration to avoid the problem of hardware and software obsolescence. An alternative approach is to record digital image information on media with recognized longevity, in a human-readable format. As an investigation of practical archiving in this way, we demonstrate the recording and retrieval of medical images and documents on to photographic film in a 35 mm microfilm format. Results indicate the level of performance achievable and the requirements for hardware and software components of such system.

Introduction

Although the idea of archiving digital image data to microfilm in a human-readable form is not *de rigueur* in today's world of encoded bits and bytes, a rationale for its consideration was put forth by Don Willis in 1992 with *A Hybrid Systems Approach to Preservation of Printed Materials*.[1] The arguments in his report catalyzed two complementary NEH-funded projects, Cornell's Digital to Microfilm Conversion[2] and Yale's Project Open Book.[3] Because the human-readable approach to digital image archiving advocated in these efforts presumes an analog rendering, it is by necessity a lossy storage technique. Despite this, unexpectedly good image quality was achieved by writing digitally born images and documents to microfilm. These efforts were aimed primarily at bi-tonal applications. While the extension of this model to continuous-tone imagery provides a variety of challenges it can also yield a number of benefits that only human-readable data preservation on to hardcopy, be it film or paper, can offer. This paper is intended to provide a glimpse of the possible paths and means to accomplish this goal.

Consistent with the previously cited work, one tends to think of digital image archiving to microfilm with respect to high contrast document images. This is an unnecessary restriction, however. Improved exposure modulation techniques, image processing, lower contrast media, intelligent image probes, XML metadata[4] insertion, and improved recorder hardware have combined to allow high quality grayscale and color image film archiving from digital image data. Both of these image types can be cast into a common image path prior to image recording and after capture. This generalized imaging path is illustrated in Fig. 1. This path is intended to apply to preservation-quality archiving. As such, 35 mm film format is implied, but not required, throughout.

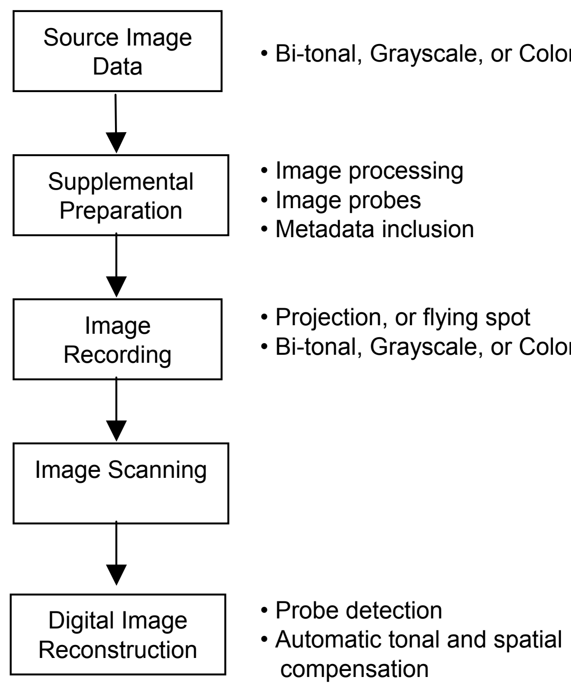

Figure 1: Imaging path for digital image archiving and retrieval with microfilm

Microfilm Archiving Image Chain

Source Image Data

There is no doubt that bi-tonal digital imagery is the most common type selected for recording to microfilm, and the number of writer choices specific to bi-tonal output reflects this popularity. So widespread is its use, one begins to wonder if practitioners realize whether digital grayscale and color output are also possibilities, which they are.

Digital image processing allows precise control of film exposures onto existing (high contrast) microfilm and can yield true preservation-quality grayscale microfilm images. For color images, one can still use monochrome films and adopt the Technicolor approach by simply recording each color record as a separate continuous-tone image. Although this approach is not truly human-readable because the color information is not integrated, the essential structure.

True color human-readable archiving has been practiced in the past using analog planetary camera imaging techniques. Both KODACHROME Film and Cibachrome have been typical film media used. From a preservation perspective, however, the results and realities of using these materials have been somewhat mixed. While some conventional color transparency materials have provided good color rendering their long term keeping properties may not suffice. On the other hand, the superior keeping properties of azo-based dyes in Cibachrome are offset by garish color reproduction and challenging chemical processing. With the improved keeping properties of traditional transparency materials, better color controls through digital processing, and high resolution color recorders these drawbacks are diminish.

Supplemental Preparation

***Pre-recording image processing*:** Many practitioners only consider post-processing image enhancement or compensation because it is only after the fact that one is faced with image degradations, which is due to recording and subsequent retrieval. Anticipating these inevitable losses and accommodating for them beforehand is actually the wiser strategy. By minimizing post-retrieval enhancements, one avoids amplifying image noise that often accompanies post-processing sharpening and contrast operators.

***Image probe inclusion*:** Just as hardcopy targets are often used to track radiometric genealogy of a source document at image capture, synthetic targets or image probes be included at the time of image recording. These elements are merely digital or synthetic versions of well-known resolution and grayscale targets often used as quality control tools, and their purpose is identical. Once recorded to film or paper, they act as a map from the archival image record back to the original digital image. They also aid in recovering the intended appearance of the image on different display devices.

These probes can also take the form of crosshair or fiducial marks and act as geometric benchmarks for image registration or geometric distortion corrections. This simple probe form can be especially beneficial because once digitized they are easily identified with automated detection algorithms. Once detected, these crosshairs then define the relative locations of tonal and resolution probes cited above. With knowledge of these locations, sharpening and tonal mappings can be automatically applied to the recovered image data. Barcodes patterns can also be added to help in administrative or quality assurance tasks.

***XML metadata*:** Digital image files are accompanied with both technical and descriptive metadata elements. elements. Any serious archiving strategy requires that these data accompany the imagery, and they can easily be recorded as human-readable text. During retrieval the characters can be recovered via optical character recognition techniques.

Image Recording

Image recorders can be separated into two types. The more popular and inexpensive of these, area projection recorders employ high quality photography of either CRT or LCD displays. An image from a small display, at the appropriate magnification, is focused onto film and photographed. The entire image frame is captured simultaneously, much the same as taking a picture of a TV screen. The other type uses a single flying spot of light or electron beam that sweeps the film in a raster fashion. The beam is modulated according to the digital signal. No projection imaging lens, *per se*, is used. These recorders tend to be more expensive and therefore less available. They often exhibit superior image quality because they are not constrained by projection optics.

Either of these recorder classes is capable of bi-tonal, grayscale, or color output. Most, however, are constrained to monochrome bi-tonal output. At first glance, the costs of developing new grayscale writing capabilities may seem an economic impediment to true grayscale image recording. But there is a software solution. By combining grayscale image processing techniques mentioned above with common digital halftoning methods, high quality, approximately continuous tone images can be rendered onto microfilm with bi-tonal recording devices.

The preferred recording choice would be a continuous tone writer, but there are very few choices for these devices. Electron beam (E-beam) recorders are the high-end performers in this category but can only write to monochrome media. This class of recorders was used by Kenney[2] and is of the single beam type mentioned above. While available, writers for recording true color (vs. color separations) imagery are few and have not achieved the high spatial resolutions of single beam monochrome-only devices. They are nevertheless very attractive choices for truly human-readable color image archiving that avoids the reconstruction processing of monochrome separations.

Image Scanning

The film retrieval portion of the image chain in Fig. 1 requires perhaps the least effort to enable accurate scanning of recorded image information. At their heart, today's microfilm scanners already have the implicit ability to capture grayscale and/or color information. The data pipelines or memory requirements may not be in place but these are not fundamental limitations to such a task. A number of high-volume film scanners within the professional photography market are up to the task for 35 mm roll film.

Digital Image Reconstruction

Although the scanned film data is in hand, it is likely to differ from the original source image data. This is where the supplemental image probes, XML metadata elements, or barcodes participate. With simple predefined locating guidelines, the reconstruction software can find the appropriate probe feature in the scanned data. Once found, software can interrogate the image features for information on how to best reconstruct the scanned image data itself into its original digital form. While the entire process is necessarily lossy, incremental improvements in recording, scanning, media, and image processing can combine to minimize these losses.

Demonstrations

We chose two practical examples as demonstrations.[5] The first addresses the recording and reconstruction of a digitally born color document using grayscale color separations. This case illustrates image and supplemental image probe recording. After scanning, these probes are interrogated for accurate tonal rendering and color channel registration of the reconstructed image.

The second example is a digital radiograph of a hand. The legal requirements for medical records often dictate that they be maintained for long periods of time, making them candidates for image archiving. This particular example was chosen to investigate whether the high quality standards associated with diagnostic medical images could be maintained through the recording, storing, and retrieval from 35mm microfilm.

Grayscale Color Separations

Given preservation longevity questions around archiving color images to color transparency or color negative films, one is frequently left with performing color separations on to film media in order to maintain color information. This can be a very unpalatable choice since the separate frames may become detached from each other and one must also deal with frame registration at some future date. Williams[5] describes one solution to these problems. By spatial integrating all three records into a single standard 35 mm frame, the separate color channels remain physically connected. An example of this is shown in Fig. 2.

In this example, recording of the color records was achieved first transforming the color image data into the CIELAB color space. The resulting L^*, a^* and b^* records were recorded as separations. This allowed the a^* and b^* color records to be spatially sub-sampled by 2 times in each direction. This is similar to the color encoding in the NTSC television, and the sub-sampling of the chroma records in JPEG compression standards. As suggested earlier, tone scale and registration marks were also added to the recorded image to aid in accurate and automatic spatial and digital image reconstruction.

Digital Radiograph Recording/Retrieval

The series of images in Fig. 3 illustrates the potential for human-readable archiving of health-related images. In this example the images were created from 12 bit/pixel digital radiographic image data, recorded onto 35mm high resolution microfilm using a 4 micron 8 bit/pixel laser beam writer. The resulting microfilm image was then scanned with a commercially available graphics arts scanner at 4000 dpi and tone corrections applied.

Despite being two generations removed from the original radiographic data, the reconstructed digital image was informally judged to be of potential diagnostic value. This was unexpected, and demonstrates what is possible using the above image quality control features, for this demanding application. It should be noted that various types of diagnostic images have different image quality requirements. These are often described in terms of statistical information measures based on a noise-equivalent patient exposure.[7] Setting the imaging requirements for the archiving and subsequent display can be approached in the same way.

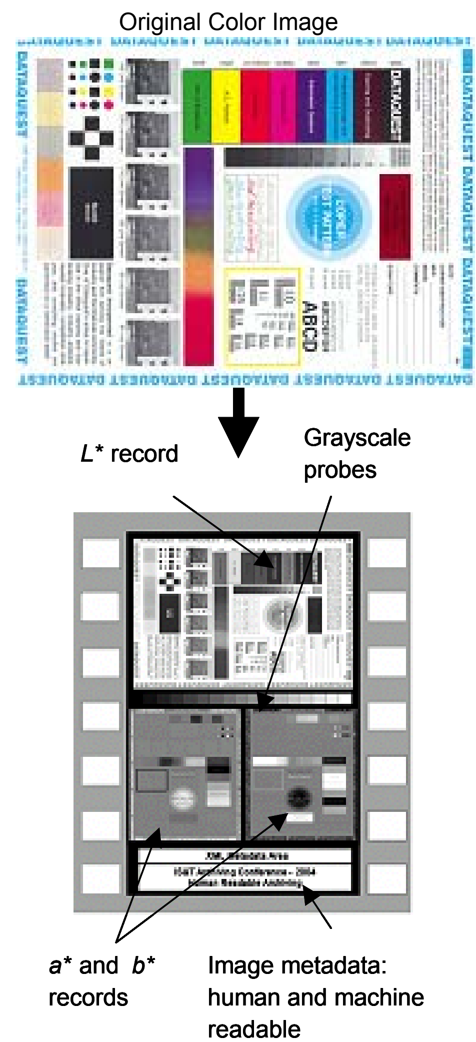

Figure 2: Color separation concept onto 35mm microfilm

Conclusions

We have described the required hardware, media and image processing path for human-readable preservation of digital images and documents to film. Two examples were chosen to demonstrate the capabilities of such a digital archiving system based on a 35mm microfilm format. The use of embedded supplemental image probes, and pre-recording image processing were instrumental in achieving these results.

Figure 3: Original radiograph (a), detail of same (b), and (c) detail of reconstruction after being recorded on and scanned from 35mm microfilm

Acknowledgments

Gratitude to Bob Breslawski, Lynn Fletcher-Heath, Bill Moore and members of the E-beam film recorder crew for providing digital data, media, and processing expertise.

References

1. D. Willis, A Hybrid Systems Approach to Preservation of Printed Materials, The Commission on Preservation and Access, 1992, http://palimpsest.stanford.edu/byauth/willis/hybrid/
2. A. Kenney, Digital to Microfilm Conversion: A Demonstration Project, Final Report to the National Endowment for the Humanities, PS-20781-94, 1996, http://www.library.cornell.edu/preservation/com/comfin.html - _Hlk396124650
3. P. Conway, Yale University Library's Project Open Book: Preliminary Research Findings, *D-lib Magazine*, Feb. 1996, http://www.dlib.org/dlib/february96/yale/02conway.html
4. E. R. Harold, Effective XML, Addison-Wesley, 2003.
5. Poster presentation at this conference.
6. D. Williams, Method for Recording and Storing Color Images, U.S. Patent #5606379, 1996.
7. I. A. Cunningham and R. Shaw, Signal-to-noise optimization of medical imaging systems, *J. Opt. Soc. Am. A*, **16:** 621-632, 1999.

Biographies

Don Williams received both B.Sc. and M.Sc. degrees in Imaging Science from RIT, and works in Electronic Imaging Products, Research and Development at Kodak. His work at Kodak focuses on quantitative signal and noise performance metrics for digital capture imaging devices and imaging system simulations. He currently co-leads the TC42/WG18 standards effort for both digital print scanner (ISO 16067-1) and digital film scanner (ISO 16067-2) resolution measurement. Mr. Williams is a frequent contributor and advisor on digitization fidelity issues for the library and museum communities. E-mail: wilyums@rochester.rr.com

Peter Burns studied Electrical and Computer Engineering at Clarkson University, receiving his B.Sc. and M.Sc. degrees. In 1997, he completed his Ph.D. in Imaging Science at Rochester Institute of Technology. After working for Xerox, he joined Eastman Kodak Company, where he works in Electronic Imaging Products, Research and Development. His technical interests include; system evaluation, simulation, and the statistical analysis of error in digital and hybrid systems. E-mail: peter.burns@kodak.com