

Cornell University Library: Preserving Cornell's Digital Image Collections: Implementing an Archival Strategy

On-the-Fly Conversion: TIFF2PNG vs. TIFF2GIF Conversion Utilities

Rich Entlich & Oya Y. Rieger

*** draft ***

Why On-the-Fly Conversion?

Document scanning that emphasizes full, high-fidelity capture has many advantages, including the creation a rich digital master that can serve many needs. One of the drawbacks, however, is the difficulty in delivering such master files via the Internet. High-resolution master files take a long time to transmit and cannot be comfortably viewed, even on state-of-the-art displays, due to their large pixel dimensions. Therefore, a central part of Cornell's digital preservation strategy has been to place the high-quality master image files at the center of the delivery systems and to develop on-the-fly conversion capabilities to create derivatives to meet varying user and staff needs. As strongly advocated by John Price-Wilkin, this approach supports flexibility in delivery, without reliance on the creation of static derivative images that would support only certain time-sensitive and specific needs. This approach also helps to reduce costs as derivatives are created only when there is a request.¹

On-the-Fly Conversion Software

On-the-fly conversion is an elegant solution to the Web delivery problem associated with high-quality master images. Made feasible in recent years by the wide availability of high-speed workstations, on-the-fly conversion can quickly produce small yet legible versions of master files on an as-needed basis. The technique validates the theory of the rich digital master by allowing it to serve multiple purposes without requiring the long-term storage of derivative files.

On-the-fly conversion has been most commonly used to scale high-resolution, bitonal TIFF files to lower resolution, gray-enhanced GIF files for Web delivery. A fast conversion utility called *tif2gif* was developed by the University of Michigan to carry out TIFF to GIF conversions and allows the selection of the degree of scaling and the number

¹ For a full discussion of the virtues of on-the-fly conversion, see: John Price-Wilkin. *Enhancing Access to Digital Image Collections: System Building and Image Processing in Moving Theory into Practice: Digital Imaging for Libraries and Archives* by Anne R. Kenney, Oya Y. Rieger. Mountain View, CA: Research Libraries Group, Inc., 2000.

of bits of gray to add. Conceivably a TIFF file could be scaled and gray-enhanced without conversion to GIF, but TIFF is not one of the Web's native graphics formats. GIF, on the other hand, has been widely supported by Web browsers since the early days of the World Wide Web. Though GIF is limited to 8-bit depth, gray enhancement of TIFF files usually employs only three to five bits of gray.

However, despite the absence of any technical limitations in using GIF as the target for on-the-fly conversion of high-resolution bitonal files, there are other good reasons to seek a substitute. The compression algorithm used in GIF files is patented, and the patent holder, Unisys Corp., has been pursuing royalty payments in recent years from software producers that use its algorithm and from web sites that employ GIFs whose origin in licensed software cannot be proven. As a result, the Internet community developed PNG (portable network graphics) a royalty-free graphics file format to take the place of GIF.

PNG is emerging, along with the already established GIF and JPEG, as the third natively-supported graphics format on the Web. As part of this investigation, we researched the status of PNG and reported the results in the FAQ of the August 15, 2000 issue of RLG DigiNews (see <http://www.rlg.org/preserv/diginews/diginews4-4.html#faq>). Given PNG's improving support and a growing movement to retire GIF as an active Web graphics format, it makes sense to examine the available tools for converting high-resolution, bitonal TIFF files to PNGs instead of GIFs.

TIFF2PNG vs. TIFF2GIF

Thus our second task was to conduct a survey of TIFF to PNG conversion tools. We started our survey close to home. We were aware that a Cornell staff member had created a TIFF to PNG conversion utility (`tif2png`) and was using it to create Web graphics for some sites. We were curious how the resulting PNG files would compare to corresponding GIF files created by the University of Michigan `tif2gif` utility. We were interested in parameters such as conversion speed, output quality, flexibility of gray level enhancement, etc.

However, in discussing the nature of the `tif2png` utility with its creator, it became evident that no significant differences should be expected. `Tif2png` was created by taking the program code from `tif2gif` and changing only the portion that converts the TIFF file to another format. The code that does the scaling and the gray-level enhancement is identical. Since GIF and PNG both use lossless compression, there should be no differences in quality.

Further discussions with the program's creator revealed that in his experience, the speed of conversion is very similar, with `tif2gif` slightly faster than `tif2png` on some files and vice versa. Converted GIF and PNG files resulting from the same starting TIFF file were closely compared and found to be identical down to the last bit, as long as the same

command line parameters (degree of scaling and number of bits of gray enhancement) are used.

We found one other TIFF to PNG utility, co-authored by one of the designers of the PNG format and available at <http://www.libpng.org/pub/png/apps/tiff2png.html>. This utility is available for the DOS and Linux operating environments, but in its present form supports neither scaling nor gray-level enhancement. Therefore, it does not serve as a suitable substitute for `tif2gif`. It is possible that this utility could be part of the toolkit for delivering gray scale and color TIFF files over the web (as converted PNG files), since `tif2gif` only works on bitonal images. However, it would be desirable to incorporate variable scaling to reduce the size and resolution of the images.

During on-the-fly conversion investigation, we discovered that the grayscale settings the `tif2gif` utility that Cornell is using was not optimized for the delivery of bitonal images with illustrations. This problem was fixed at once.

The third leg of our investigation was to query the digital library community about its current level of PNG utilization and for any observations about the quality or usability of those images. A message² was sent out on several listservs requesting feedback. Only two substantive replies were received.

One institution reported that upon receiving a request for an image, if its web site detected a PNG-enabled browser, it would deliver the image as a PNG instead of as a GIF. Another institution reported that it had considered using PNG for image delivery and had hired a contractor to run some tests. The images received were judged to be unacceptably distorted for their intended purpose. Given that PNG is a lossless raster image format, there is no reason files converted to PNG from other raster formats to show any distortion at all. So the problem reported probably resulted either from the use of improperly written conversion tools or errors on the part of the contractor. Either way, the incident points out some of the difficulties faced by a new format trying to gain acceptance against established formats, even when it offers superior functionality and the absence of licensing difficulties.

² Cornell University Library, Department of Preservation and Conservation is involved in a small project to compare the quality and utility of GIF and PNG file formats in the delivery of access image files. The goal of the study is to compare various attributes of GIF and PNG file formats for on-the-fly delivery purposes (generation of GIF or PNG files for online delivery from master TIF images when there is a request). Although the PNG file format has gotten quite a bit of publicity (and partial Explorer and Navigator support) during the last couple of months, it looks like the file format has not yet become very popular in image delivery on the Web. Please contact me if you know of any institutions or projects that use PNG file format for master or access images.

Conclusion

In summary, our investigation revealed that although PNG is a well-designed replacement for GIF, with numerous advantages, it currently faces several obstacles which preclude its immediate adoption as a target for delivery of converted, high-resolution, bitonal TIFF files. These obstacles include insufficient web browser support and an overall lack of knowledge and interest on the part of the library imaging community.

The on-the-fly conversion study was limited to bi-tonal images. We have decided to continue using the tiff2gif conversion software (for delivery of bitonal text-based images) for the time being. However, we will continue our investigation as new file formats and compression techniques continue to emerge. One interesting trend that we are closely watching is the increasing popularity of file formats that provide nested-resolution (a.k.a. tiling formats). This category of file formats (e.g., FlashPix, GridPix, and JTIP, JPEG2000) are particularly appropriate for tonal images, and have the advantage of fast delivery and high quality. Also certain compression techniques such as Wavelet allow panning and zooming to fit the file to the purpose on-the-fly. With server-side wavelet compression, users can dynamically create JPEG derivatives at various resolutions and focus on specific segments of an image.