

Digitization of Photographic Collections

Franziska Frey
Image Permanence Institute, Rochester Institute of Technology
Rochester, New York

Abstract

Almost every major institution is contemplating some kind of digital-imaging project, but sources of information and experience are few. Technical aspects of digital imaging present libraries and archives with very difficult and complex choices. There are as yet no codified technical standards for image capture, display, and output, all of which affect the image quality, the cost, and, ultimately, the success or failure of the entire undertaking. Even if a vendor will provide a finished "turnkey" system, an institution must understand the nature of the digital-imaging product they are buying. They must know how much image quality and functionality can be expected from it, both now and in the future. New tools must be provided for collection managers to make that possible. Beyond these purely technical issues, institutions must be able to relate the digital-image database system to the fundamental collection activities of access and preservation.

Despite all the possibilities for manipulating digital images, image quality choices made when files are first created have the same "finality" that they have in conventional photography. They will have a profound effect on project cost and the value of the final project to the users. Image quality requirements therefore have to be established carefully before a digitization project starts.

Introduction

There are no guidelines or accepted standards for determining the level of image quality required in the creation of digital-image databases for access or preservation of photographic collections. As a result, many institutions are already disappointed because their efforts don't lead to the results they were hoping for. Either the parameters chosen for the digitization were not thought through, or technology has already changed since the project started—leaving people disappointed with the digital images they have produced. However, nobody knows what technology will be available in a few years, and choosing the right scanning parameters is a task still needing to be researched. One problem is that, right now, the cycle of understanding image quality has only started for the new imaging technologies. Some of the failures could have been prevented; some of them are a result of the fast technology change.

One of the big issues is that institutions will have to decide beforehand on the use of their digital images. This still causes a lot of questions and problems. Sometimes the use of the images is not clear when a project starts, but more often institutions don't take enough time to think through the use of the digital images. Will the images be used on an image data base on the World Wide Web? Will they be used for printing of reproductions? What size will the prints be? These are only a few questions that have to be answered *before* a digitization project starts.[1]

Moreover, it has to be kept in mind that scanning from an archive is different than scanning for prepress purposes. As seen at different conferences, this is a new concept for both the museums and the technical field, and it will need some work to make people understand where the problems are.

Image Quality Framework

Working on different aspects of image quality has shown us that the parameters are difficult to define and that they have to be looked at as a whole "quality framework." The more one looks at image quality and possibilities to clearly define it, the more parameters have to be taken into account. When looking at image quality, the whole image processing chain has to be examined. Besides issues concerning the scanning system, compression, file formats, image processing for various usage, and system calibration are some of the areas to be looked at closer.

Advances in image-data compression and storage-media development have helped to reduce the concern about storage space of large data files. Nevertheless, image compression in an archival environment has to be evaluated very carefully. Because, in this case, the future use of a digital image is not yet determined, one copy of every image should be compressed using a lossless compression scheme. Lossless compression makes it possible to exactly reproduce the original image file from a compressed file. New compression schemes, like wavelets, which do not produce the well-known artifacts that JPEG compressed files show, are still not readily available. It is not yet clear which of the schemes will be the preferred method.

New file formats will be introduced. The files also need to be in a format that can be read and retrieved in the future. It is advisable to use standardized formats where source

coding is readily available and/or in public domain for image data bases.

Image quality is affected by the sequence of applying different image processing steps including compression. It is important to be aware of the effects of different processing algorithms, and here also the unknown future use of the images has to be taken into account. Image processing done before storing the images can affect the quality of future processing. Therefore it is, for example, recommended not to sharpen the highest quality file before storing.

A common problem is the difference between the images when viewed on the various systems/monitors in a given working environment. Calibration of all the systems involved in the imaging project should be required. The situation gets more complicated if originals and images on the screen are viewed side-by-side, because in this case the observers are not allowed to adapt to each "environment" individually.

Digitizing System

Successfully digitizing a photographic collection requires as much experience as conventional reformatting. The ease of use of many digitizing systems has fostered the perception that scanning is "simple." Additionally, most of the available scanning technology is still based on the model of immediate output on a current output device with the original on hand during the reproduction process. Spatial resolution and color mapping are determined by the intended output device. Depending on the quality criteria of the project, a more sophisticated system and more expertise by the operator are needed to successfully digitize a collection in an archival environment where the future output device is not yet known. The characteristics of scanning devices such as optical resolution, dynamic range, registration of color channels, bit-depth, noise characteristics, and quantization controls need to be carefully evaluated with consideration of the final use of the digital images.[2, 3, 4]

Furthermore, when choosing the digitizing system, it has to be kept in mind that approaches that work for a small number of images might not be suitable for the large number of images usually found in collections.

The image quality framework should include test targets for objective testing of the four main image quality parameters:

- Tone reproduction
- Detail and edge reproduction (MTF)
- Noise
- Color

Discussions with various people in the field have shown that it is important to emphasize that *targets are about the scanning system and not about collections.* This means that the tests to be performed are primarily aimed at characterizing the scanning systems.

At this point in time, scanning the actual photographs will still need the intervention of a well-trained operator. In a few years, some of these tasks will be automated and manual interventions will be less and less necessary.[5] One could say that targets will then be about collections because no matter what original is scanned, it will automatically turn out right.

Each one of the main image quality parameters needs special targets for the different forms of images (e.g., prints, transparencies, etc.). The targets should consist of the same material as the materials that will be scanned—photographic paper and film.

Full versions of targets could be scanned every few hundred images and then linked to specific batches of production files, or smaller versions of the targets could be included with every image.

In addition, the spectral sensitivities of the scanning system should be known, and a complete description of the image processing chain must be at hand.

After targets are scanned they are evaluated with a software program. Having an objective tool to compare different scanning devices will be more and more important. Up to now scanner manufacturers usually have used their own software when evaluating and testing systems. Some of the information coming out of these tests and additional information, like spectral sensitivities and details about the actual image processing chain, will be used to characterize the scanning system. Other information will be associated with the image files; it will be used to perform some data corrections later on as the images are being processed for output or viewing. Standardized approaches and data forms are required for interchangeability of the data. To quote M. Ester:[6]

> If I see shortcomings in what we are doing in documenting images, they are traceable to the lack of standards in this area. We have responded to a practical need in our work, and have settled on the information we believe is important to record about production and the resulting image resource. These recording procedures have become stable over time, but the data would become even more valuable if there was broad community consensus on a preferred framework. Compatibility of image data from multiple sources and the potential to develop software around access to a common framework would be some of the advantages.

In addition to these objective measurements of image quality, taking into account subjective judgments from users allows the inclusion of aesthetic issues in the final quality decisions.[7] This is especially important because human judgment decides the final acceptance of an image.

Quality Control

The best approach to digital image quality control includes, on one hand, subjective visual inspection and, on the other hand, objective measurements performed in software on the digital files themselves.[8] Efforts should be made to standardize the procedures and equipment for subjective evaluations by means of monitor and printer calibration. For objective image quality measurement, software should be available which is designed to locate and evaluate specific targets and then report numbers or graphs describing key image quality parameters. Such software should ideally be a plug-in to a full-featured image browser so that the review of all aspects of the image file (header info, index, and tracking data, etc.) can be reviewed at one time.

A key point is that targets and the software to evaluate them are not just for vendor checking—they serve to guarantee the long-term usefulness of the digital files and to protect the investments of the institution.

Conclusions

The many advantages of the emerging digital technologies for photographic collections are obvious, but there is still a long way to go. The technology is still young and changing rapidly. This is the reason why museums, libraries, and archives that want to use digital imaging have to buy the know-how or find suitable partners.

In the long run, the required know-how should be available on site in the collection. Proper custodianship of collections must now employ a greater awareness of disciplines previously far removed from the established practices of the photographic archives. Intellectual property law, high-speed data transfer technology, and database management are but a few of the specialties demanding the attention of collection managers.

Most importantly, communication has to be improved among all the participating parties. Due to the fast-changing and complex imaging technologies involved, collection managers need to work together with engineers and imaging scientists, who often lack collection-related knowledge. Both sides need to be willing to learn the special problems and needs of the other party.

High-quality digital image archiving is neither an easy nor a cheap endeavor. “High-quality” is not only defined by high information content of digital files, but also by the viability of access and data integrity over time. For most collections, the cost factor will ultimately limit the digital image quality they can achieve and consequently the use of the files. However, preserving collections through conventional reformatting is not cheap either. The creation of large digital image collections is not likely to be attempted more than once a generation. This means that it had better be done right the first time, so being aware of the technical nature of the digital images produced is quite important.

References

1. F. Frey and S. Süsstrunk, “Image Quality Requirements for the Digitization of Photographic Collections,” *Very High Resolution and Quality Imaging*, SPIE Vol. 2663, 1996.
2. J. Holm, “Survey of Developing Electronic Photography Standards,” *Critical Reviews of Optical Science and Technology*, SPIE Vol. CR61, pp. 120-152, 1996.
3. S. E. Reichenbach et al., “Characterizing Digital Image Acquisition Devices,” *Optical Engineering*, Vol. 30 (2), pp. 170-177, 1991.
4. ANSI/AIIM TR26-1993, *Resolution as it Relates to Photographic and Electronic Imaging*, Copyright 1993 by Association for Information and Image Management, 1100 Wayne Avenue, Suite 1100, Silver Spring, MD 20910-5699, 1993.
5. J. Holm, “Factors to Consider in Pictorial Digital Image Processing,” *Proceedings IS&T’s 49th Annual Conference*, Minneapolis, MN, May, pp. 298-304, 1996.
6. M. Ester, *Digital Image Collections: Issues and Practice,* Commission on Preservation and Access, 1996.
7. G.A. Gescheider, *Psychophysics—Method, Theory, and Application*, 2nd Edition (London, Hillsdale, NJ: Lawrence Erlbaum Associates), 1985.
8. Y. Kipman, “Image Quality Metrics for Digital Image Capture Devices,” *Proceedings IS&T’s 48th Annual Conference*, Washington, DC, pp. 456-459, 1995.