mcn

A PUBLICATION OF THE MUSEUM COMPUTER NETWORK

ISSN 1042-3729

GUEST EDITOR
Günter Waibel

INTERNET SIG CHAIR
David Bridge

STANDARDS SIG CHAIR
Rob Lancefield

SMALL MUSEUMS SIG CHAIR
Annita Andrick

DIGITAL IMAGING SIG CHAIR
Günter Waibel

CONTROLLED VOCABULARY SIG CHAIR
Leah Prescott

ARCHAEOLOGY & ETHOLOGY SIG CHAIR
Kevin DeVorsey

INTELLECTUAL PROPERTY SIG CHAIR
Amalyah Keshet

CALIFORNIA SIG CHAIR
Marla Misunas

EDUCATION SIG CHAIR
John Chadwick

LIBRARIES SIG CHAIR
Cameron Trowbridge

SPECTRA DESIGN
Christopher Seki

PRODUCTION MANAGER
Kathleen Buckner

MCN LOGO
Fumiko Toyosumi

PRINTER
West America Graphics

MANAGEMENT SERVICES
Provided by Adler Droz Inc.
1550 S. Coast Hwy, Suite 201
Laguna Beach, CA 92651

Spectra reports on museum information management and applications of computers in museums. Articles, news items and letters to the editor are welcome. Materials may be submitted via e-mail or on disk saved as Word 6.0 or text files. Deadlines for submissions are the 1st of May, August, and November.

The editorial staff reserves the right to review, select, delete and correct copy submitted for publication. Authors assume full responsibility for the accuracy of their statements and for the opinions of judgements that they may express. Those opinions do not necessarily represent the views of the Museum Computer Network.

The Museum Computer Network does not formally endorse or recommend any of the publications, products, services, programs or organizations which may be mentioned in *Spectra*, and assumes no liability for the accuracy of the information reported to us by outside sources

Reproduction is permitted provided that reference is made to *Spectra*, the title and author of the specific item, and the date of issue.

Send Correspondence regarding
Spectra to:

MCN
Attn: Spectra Editor
1550 S. Coast Hwy, Suite 201
Laguna Beach, California 92651
877.MCN.3800 • 949.376.3456 (fax)
e-mail: spectra@mcn.edu

Museum Computer Network

Board of Directors

President
Susan Patterson
St. Louis Art Museum
1 Fine Arts Drive
St. Louis, MO 63110
314-721-0072 x278
314-721-6172 (fax)
spatter@slam.org

Vice President/President-Elect
Leonard Steinbach
Chief Information Officer
The Cleveland Museum of Art
11150 East Boulevard
Cleveland Ohio 44106-1797
216 421 7340 ext 306
steinbac@cma-oh.org

Treasurer
Charles Patch
Director of Systems
The Historic New Orleans Collection
533 Royal Street New Orleans, LA 70130
504-598-7116
504-598-7108 (fax)
chuckp@hnoc.org

Secretary
Greg Spurgeon
National Gallery of Canada
380 Sussex Drive
P.O. Box 427, Station A
Ottawa, Ontario
Canada K1N 9N4
613-990-0482
613-990-8689 (fax)
gspurgeo@gallery.ca

Jim Angus
Manager of IT and Web Development
Natural History Museum
900 Exposition Blvd.
Los Angeles, CA 90007
213-763-3317
213-749-4308 (fax)
jangus@ed-resources.net

Murtha Baca, Ph.D.
Head, Standards
Getty Research Institute
1200 Getty Center Drive, Suite 1100
Los Angeles, CA 90049
310-440-6339
310-440-7780 (fax)
mbaca@getty.edu

Jim Devine
Head of Education and Digital
Media Resources
Hunterian Museum and Art Gallery
University of Glasgow
Glasgow G12 8QQ Scotland
++ 44 (0)141-330-3691/4221/2689
++ 44 (0)141-330-3617 (fax)
jdevine@museum.gla.ac.uk
Web: http://www.hunterian.gla.ac.uk
http://www.gla.ac.uk/museum/people/jdevine/jim.html

Robin Dowden
Web Site Manager
Walker Art Center
725 Vineland Place
Minneapolis, MN 55403
612-375-7541
612-375-7575 (fax)
robin.dowden@walkerart.org

Amalyah Keshet
The Israel Museum, Jerusalem
P. O. B. 71117
Jerusalem, Israel 91710
+972-2-670-8874
+972-2-670-8064
akeshet@netvision.net.il

Marla Misunas
Collections Database Manager
San Francisco Museum of Modern Art
151 Third Street
San Francisco, CA 94103-3159
415-357-4186
415-357-4158 (fax)
mmisunas@sfmoma.org

Sam Quigley
Director of Collections
San Francisco Museum of Modern Art
151 Third Street
San Francisco, CA 94103-3159
415-357-4061
squigley@sfmoma.org

Richard Rinehart
Berkeley Art Museum/Pacific Film Archive
University of California
Berkeley, CA 94720-2250
510-642-5240
510-642-4889 (fax)
rinehart@uclink2.berkeley.edu

CONTENTS

EDITORIAL
TO PRESERVE OR TO PRESENT
Günter Waibel
6

REVEALING HISTORY - DIGITAL IMAGING THE NEW PHOTOGRAPHIC RESEARCH TOOL
Jill Marie Koelling
10

BUILDING A DIGITAL ARCHIVE - DIGITIZING FOR THE LONG TERM
Carol Hernandez and Robin Lilien
16

WHAT ARE DIGITAL IMAGE MANAGERS?
Dr. S. K. Hastings
18

LIGHT EXPOSURE TO SENSITIVE ARTWORKS DURING DIGITAL PHOTOGRAPHY
Ben Blackwell
24

CONSERVATION ISSUES IN DIGITAL IMAGING
Dr. Norbert Lossau, Martin Liebetruth
30

PRODUCE, PUBLISH AND PRESERVE: A HOLISTIC APPROACH TO DIGITAL ASSETS MANAGEMENT
Günter Waibel
38

THE ALBERT M. GREENFIELD DIGITAL IMAGING CENTER FOR COLLECTIONS
ACADEMY OF NATURAL SCIENCES, PHILADELPHIA
Eileen C. Mathias
46

PRACTICAL APPROACH TO COLOR MANAGEMENT
Andrew Stawowczyk Long
50

DO YOU SEE WHAT I SEE?
Trudy Levy
56

ISSUES IN DIGITAL ARCHIVING
Bruce Fraser
60

CONTENT-BASED IMAGE RETRIEVAL: A NEW PARADIGM FOR IMAGE RETRIEVAL?
Colin C. Venters and Dr. Matthew D. Cooper
66

To Preserve or to Present?

Museums pursue two main objectives: to preserve and to present cultural heritage materials. Trying to faithfully serve both imperatives of their charge often results in a dilemma. The need for preservation and presentation clash over the wear and tear involved in all forms of providing access to collections. Exhibition time reduces the lifetime of an object, a basic fact museums have had to contend with ever since they embarked on their mission as the guardians of human artifacts.

Digital imaging presents the old dilemma in a new garb. Digitizing a collection means exposing it to handling and light, both of which impact the lifespan of an object. However, digital imaging also presents us with the opportunity to fulfill the imperative of access in unprecedented breadth and depth. To put it country simple, by digitizing their artifacts, museums can show more of their collections to more people all over the world. The Internet obliterates the physical limitations of gallery space and museum location.

I think the questions pondered and the answers provided by the contributors to this special issue of *Spectra* on digital imaging show a way of sidestepping the dilemma. Rather then conflict, preservation and presentation strategies may actually re-enforce each other. The California Digital Library presumes that by following its procedures and specifications for digitizing objects, an institution can build an archive of digital master files with an estimated lifespan of about 50 years. While that figure seems optimistic from my perspective, it still points in the right direction: digital images provide a return far beyond the short term.

A well-planned and executed digital imaging project has the potential to provide access to collections for decades to come. According to Ben Blackwell (please see his article on “Light Exposure to Sensitive Artworks during Digital Capture” for more details), a high-end digital capture exposes an artifact to the same amounts of light as two days under average gallery lighting conditions. Let’s assume for the sake of argument that the handling involved in exhibiting and digitizing inflicts the same amount of stress on an object, thus rendering it negligible for our comparative purposes. Now, if we think of our dilemma as a trade of “days of object lifetime” for "public access," we just bought virtually unlimited access over a period of decades for the price tag of a two-day spatially limited exhibition. Sounds like a bargain to me.

Furthermore, the availability of accurate digital surrogates cuts down on the handling of original objects by both staff and researchers; the digital images answer most needs. (As Jill Marie Koelling’s piece illustrates, in some cases the digital surrogate provides even more answers than the original did.) Instead of stealing from the object’s lifespan, the digital image now gives back by retiring the original to its ideal storage environment for an increased period of time.

The mutual re-enforcement of preservation and presentation in digital imaging hinges on two interrelated elements: the quality of our digital images, and the quality of our care for them. Both factors determine the longevity of our data or, in other words, the return for our efforts. Digital images in and of themselves present an entirely new subset of preservation issues revolving around the question of how to ensure the permanence of digital information in an ever-changing technological environment. Naturally, the range of issues exceeds the scope of this publication. However, the essays gathered here should help you make informed decisions or, at the very least, prompt you to ask the right questions about critical aspects of your digital endeavor such as direct capture, color management and digital asset management. As an advanced stage of preserving the artifact, now it has become time to learn how to produce, manage and preserve its digital representation.

I hope you'll enjoy this issue of *Spectra*. To learn more about digital imaging, subscribe to the Digital Imaging SIG's e-mail list on the MCN Web site, or stop by the SIG meeting at the annual conference. I'll see you then!

Günter Waibel
Digital Media Developer
Berkeley Art Museum & Pacific Film Archive
Digital Imaging SIG Chair

NEWS FLASH

MARK YOUR CALENDARS NOW!

CIMI/MCN 2001 Annual Conference

Real Life: Virtual Experiences

New Connections for Museum Visitors

Cincinnati, Ohio • October 24-27, 2001

ABOUT CIMI/MCN 2001

For the year 2001, the Museum Computer Network's annual meeting will be a special event. MCN is partnering with the CIMI Institute—the leading provider of training in digital museum applications—to host a three-day exploration of how new technologies can and will be used to affect visitors' experiences in the museum environment. Topics to be addressed include technical and infrastructure issues, business strategies, and educational and social implications of new technologies impacting the museum visitors' experience.

ABOUT CINCINNATI

Cincinnati is a magnificent city surrounded by hills on the Ohio River. Numerous bridges connect the city to its suburbs and beautiful, rolling green hills of Kentucky. From turreted German mansions and Federal-style homes to contemporary skyscrapers, it is a city of diversified landscapes. It is only a short distance from the Westin to an extraordinary new aquarium, a live theater, nightclubs, fine dining, shopping, and several world-class museums.

PLANS ARE ALREADY UNDERWAY TO INCLUDE EVENTS AT ONE OR MORE OF CINCINNATI'S MUSEUMS AS PART OF THE CONFERENCE PROGRAM. LOOK FOR MORE INFORMATION ABOUT THE SIGNIFICANT AND INFORMATIVE EDUCATIONAL SESSIONS IN UPCOMING CONFERENCE NEWS FLASHES.

FOR MORE INFORMATION, VISIT THE MCN WEB SITE AT WWW.MCN.EDU
OR CALL THE MCN OFFICE AT (877) 626-3800.

CIMI / MCN 2

CALL FOR PAP

For the year 2001, the Museum Computer Network's annual meeting
to host a three-day exploration of how new technologies can and w
ment. The goal of this conference is to create environments—throug
pants can engage each other and learn methodologies they can take
and how it can be done. The event will be organized around several t
to enhance visitors' experiences.

Whether hand-held devices that visitors use while touring an exhibitio
designed to enhance pre- or post visitor experiences, museums are
experience that people have when visiting our physical spaces. CIMI/
and is imagined in the future—and the technological possibilities for

You are invited to participate in CIMI/MCN 2001 by submitting a proposal fo
be reviewed by a program committee to ensure a tightly focused, high quality c
and the preliminary program will be announced on June 15, 2001.

Selected conference proceedings will be published in the winter issu
conference proceedings must have full papers to MCN by **AUGUST**

CIMI/MCN 2001 CONFERENCE THEMES

Technology Affecting the Visitors' Experience
- Hand-held and palm devices
- Learning centers and kiosks
- Personal information assistants
- Pre- and post-visit websites
- Responsive public spaces and ubiquitous computing
- Remote networked environments
- Virtual environments

Business Strategies
- Funding innovative applications of technology
- Developing a business plan for new implementations
- Establishing collaborations and multi-institutional ventures
- Figuring the real return on your investment

Evaluation
- Matching specific technologies with visitor groups
- Who is the visitor that you are aiming to reach?
- How do we analyze the visitor's experience?

Social Implications
- Redefining the museum's relationship with visitors
- Experience planning
- Interface design
- Non-computer interfaces
- Preserving the visitor's experience in environments with competing technologies

Infrastructure and Technical Questions
- Technical infrastructures in new and old buildings
- Production facility issues
- Enriching the visitor's experience through distributed networks
- Systems integration
- Streaming media
- Using portable devices to access archives and multimedia

2001

Real Life: Virtual Experiences, New Connections for Musuem Visitors
Cincinnati, OH, USA • October 24-27, 2001

R T I C I P A T I O N

g will be a special event. MCN is partnering with the CIMI Institute will be used to affect visitors' experiences in the museum environ-ugh workshops, plenary sessions and roundtables—where partici-e home to their own museum. It's a chance to learn what is possible tracks that will focus intensively on aspects of applying technology

ion, in-gallery kiosks, "smart architecture," or web-based resources investigating technological solutions that affect the quality of the I/MCN 2001 will examine site-based computing—as it exists today or enhancing and extending the museum experience.

or a paper or workshop that addresses the conference theme. Proposals will conference program. The call for participation closes on **APRIL 30, 2001**

sue of Spectra. Presenters who want to be included in the Spectra **T 15, 2001**.

SUBMISSION REQUIREMENTS

Presentation and session proposals:

- Full identification of the speaker(s) or presenter(s), including name, job title, institution, address, phone, fax, and e-mail address
- Title of the proposed presentation or session
- Abstract describing the relevance of the topic and content to be covered; panel proposal should indicate the subjects to be addressed by individual speakers
- If submitting a panel proposal, please indicate whether panelists participation is confirmed or proposed
- AV requirements for presentation

Workshops:

- Full identification of the instructor(s), including name, job title, institution, address, phone, fax, and email address
- Workshop Title
- Outline of content to be presented and learning objectives
- AV requirements for presentation
- Note: instructors are entitled to free conference registration and $500 honorarium (full-day workshop) or 1/2 conference registration and $250 honorarium (half-day workshop)

CIMI/MCN 2001 proposals may be submitted online at http://www.mcn.edu/mcn2001/
(ONLINE SUBMITTALS STILL UNDER CONSTRUCTION)

Alternatively, proposals should be directed to:

Robin Dowden
MCN 2001 Conference Co-Chair
Walker Art Center
Vineland Place
Minneapolis, MN, 55419
voice 612.375.7541 • fax 612.375.7575
email: robin.dowden@walkerart.org

CIMI CONSORTIUM

Revealing History

Digital Imaging, the New Photographic Research Tool

Jill Marie Koelling, Curator of Photographs
Nebraska State Historical Society

Digital imaging is no longer a new phenomenon in the museum and archive community. Many institutions are actively involved in scanning their collections. The Internet, listserves and numerous publications offer abundant information about how to scan, what equipment to use, what metadata to record, the type of media to use and the many ways in which to access image files. What has not yet been addressed are the significant advantages to historic photograph research when using high-resolution digital surrogates.

Two years ago, the Nebraska State Historical Society began to generate digital image files from the Solomon D. Butcher glass plate negative collection. This project, funded by Ameritech and the Library of Congress National Digital Library Award Program, produced in electronic form the entire Butcher photograph collection of more than 3,000 images and the 1862 to 1911 letters of the Uriah Oblinger family. The Oblinger letters offer important testimony about the life of homesteaders on the Great Plains and are a wonderful extension to the images of homesteaders captured in the photographs by Solomon D. Butcher. The collections will soon be online as part of the American Memory portion of the Library of Congress Web site.

Scan from the original glass plate negative of the Pacific Hotel in Callaway, Nebraska. Photograph by Solomon D. Butcher, 1889.

Scan from a printing-out-paper print made from the original glassplate negative of the Pacific Hotel in Callaway, Nebraska. Photograph by Solomon D. Butcher, 1889.

Background

Digital imaging is everywhere. The speed at which this technology has been incorporated into our culture rivals the invention of photography itself. Movies, advertisements and news media all use some form of digital imaging. Consumers no longer exclaim over the seemingly magical ads plastering visual media. It is now cheaper than ever before to create digital image files. But deciding to create those files is just the beginning of a long list of questions to address.

Before embarking on the Butcher-Oblinger digital project, staff at the Nebraska State Historical Society identified some of the issues we felt most impacted our decision to begin scanning our collections.

There are two main reasons for getting into the digital arena. The first and most obvious is access. By creating digital image files we enable the use of collections via electronic media. More access equals more patrons served. By creating a Web site and posting images from the collections on that site, people around the world have instant access, without ever having to travel to Nebraska.

The main argument against digitizing the collections was also increased access. Some believe increased access on the Web may cause increased use of the originals. However, it is possible to retire or limit access to the original if you generate a high-resolution image file. Although too large to make available online, a high-resolution image file can be burned to CD-ROM and loaned or sold to researchers. It will be interesting to track the requests for access to original Butcher images and Oblinger letters once the collection is mounted on the Library of Congress American Memory Web site.

Preservation is the other reason to start digitizing collections. Let me make perfectly clear, however, that the Nebraska State Historical Society does not mean the digital image is a replacement for the original. It is not possible to exactly replicate the original electronically. What we can do, however, is stop the loss of information due to the ravages of time. As we all know, part of our job is preserving our collections. No matter what we do, objects continue to deteriorate. Creating a digital image of an object can freeze that object in time. On page 10 is an example of a Butcher glass plate that is broken in three places. You can see on the scan that the pieces were put together on the scanner bed and a file was generated, but there is emulsion missing from the roof near one of the breaks. The information from that portion of the photograph is gone forever. Luckily, we also have in the collection a printing-out-paper print made from the plate prior to the break. This is an unusual situation, we have the information in another form, but that is not the case for any other Butcher plate that is broken. Granted, the information on the roof area is not exceptionally significant, but just imagine if the loss occurred on the porch of the hotel.

Creating a copy photograph of the object can also record the state of an object in time, however, the results will be severely limited by the film stock. For example, modern photographic films and papers cannot reproduce all of the information found in a 19th century glass-plate negative. Digitizing that same glass plate will capture much more of the tonal range than any of today's films. The other advantage is having as your digital file what for all intents and purposes is a poster-sized print of the original as opposed to a 4x5 transparency or, worse yet, a 35mm transparency or copy print from a copy negative.

It is important to mention at this point that digitizing collections is very invasive. The Butcher glass-plate negatives had been used sparingly over the years, primarily to produce prints for John Carter's book, *Solomon D. Butcher: Photographing the American Dream*. Many of the negatives had undoubtedly not been touched since Butcher came to the Society in 1915 to work on the collection. Every negative was cleaned, scanned, re-sleeved and stored in new boxes. An added benefit to scanning the collection was our ability to pay for cleaning and re-housing the collection at the same time, but I must stress that every negative was handled. This is not usually the case in normal access situations.

The next question to answer was where to start. We all know it is impossible to scan everything. Even if you had the time and money it doesn't necessarily mean you should. So the next question becomes, how do we decide what to scan? Collection surveys are a great place. The most important component in any collection survey is demand. How often is a particular collection requested by your patrons? High-demand collections should be digitized first. The other important component is subject matter. This is how you find those collections important for content, but not frequently requested. Most collection surveys also include categories for extremely fragile materials that are in danger of rapid deterioration. These are essential to scan first, if your capture mechanism will not harm the object.

The next decisions to make were related to the technical issues of digital imaging. The five most important technical issues are resolution, file format storage, refreshment and copyright.

Resolution is the number of pixels per inch (ppi) in a digital image file. The key factor in determining resolution is final use of the file. For the Butcher-Oblinger project, the ultimate goal was Web site use; however, we felt strongly from the beginning that we only wanted to scan the collections once and therefore did not want to create image files that were only useful on the Web. Regardless of the image file resolution specified during digital capture, all images display at 72 ppi as a direct result of current monitor resolutions. The larger resolution files generally require the user to scroll across the screen in order to see the entire image. Most image files posted on Web sites were scanned at a resolution between 72 ppi and 300 ppi. The higher the resolution, the larger the file, the longer the download time.

Image files used for publications vary depending on the dot per inch (dpi) used by the publisher. A pixel and a dot are not the same thing. The computer world adopted the term dot to refer to a pixel, but it can be confusing and cause problems when dealing with publishers. For every dot in the print world, there need to be two pixels in your image file. If you are expecting to use image files for coffee-table book publications, you need at least 600 ppi of resolution. It is important to think about this when determining the resolution at which to capture your initial scan.

Is it worthwhile to scan at a higher resolution than you think you might need for immediate use? Higher resolution scans, although more time-consuming to produce and larger to store, will be useful longer into the future. Scan once, scan right and scan high. We decided to scan the Butcher images and the Oblinger letters at 800 ppi. This will ensure the longevity of the image files. Some of the Butcher negatives were scanned at 3,175 ppi, the highest optical resolution our scanner can produce. These very high-resolution files were generated because the negatives are considered the icons of the collection and we do not want to ever have to pull them again.

Jpgs? Tifs? Gifs? Once again the end use determines the format. Gifs and jpgs are used online. Thumbnails are currently going up on the Internet as gifs or jpgs, and the full-screen image files are generally jpgs. Tifs, however, are the best format to use for long-term storage of high-resolution image files. Although jpgs offer several levels of compression and it is possible to compress without losing information, we decided to save the master file, the original scan file, as a cross-platform compatible uncompressed tif.

Storage can be a big problem. Memory continues to get cheaper and bigger, but it is still a problem. Do you need fast access for everyone in your institution to the high-resolution image files? Is it possible to get away with posting jpgs on your LAN for staff and patron access, while keeping the tifs on CDs in storage? An 800 ppi RGB uncompressed tif file from an 8x10 original photographic print is approximately 160MB. The file size not only requires a large storage media, but it also requires substantial RAM memory on your computer in order to work with the file. Image file sizes are getting larger and larger. The Digital Imaging Laboratory at the Nebraska State Historical Society has generated image file sizes as large as 640 MB from an 8x10 black-and-white negative at 2,800 ppi. Computer memory is keeping up, but your institution may not be able to afford the newest, fastest PCs. Don't let that discourage you, however, from planning for the future with your choice of resolution and file format.

The most important issue to understand before embarking on a digital imaging project of any kind is refreshment. Very simply, refreshment is the transfer of files from one media to another on a regular basis to ensure no loss of information. The media is not as unstable as the computers on our desks, which are unlikely to be the same ones five years from now. It is important to update, transfer or refresh data in order to avoid a situation where you no longer have a drive that will read the disk on which the files are stored. Do you have any 5 1/4 inch disks and no 5 1/4 inch drive? Five years is the longest time to wait before refreshing data. Three years is more advisable, simply because every year computer technology seems to advance more rapidly. All files generated by the Digital Imaging Laboratory at the Nebraska State Historical Society in 1998 will be transferred to new media in January 2001. Hopefully, we will be moving files from CD-ROM to DVD.

Copyright is, of course, the most difficult issue. If there is any doubt that the material you want to scan is not in the public domain or your institution does not have the right to copy it, a decision must be made. Creating a digital file is creating a copy. Posting on the Internet is a global posting, so international copyright laws apply. There is good information about copyright available online, and you can start at the U.S. Copyright Office Web site, *www.loc.gov/copyright/*. It is probably a good idea to consult a copyright attorney if there are any doubts.

The final issue we discussed before generating our first scan was metadata. We knew it was important to generate data about the image files, and we spent quite a bit of time searching the Internet and reading about what might or might not be important. In the end, our metadata database was created from recommendations presented by Anne Gilliland-Swetland at the Northeast Document Conservation Center's School for Scanning in Berkeley, California, in May 1997. We track metadata on the following items: linking numbers between the original object and the scan file name; data regarding when the object was scanned; what scanner was used; who created the scan; what file format was chosen; what resolution was chosen; and whether the scan was grayscale or RGB. We also track metadata about surrogate files.

The Project

The Butcher collection was chosen for scanning because it depicts, like no other photograph collection, the story of European settlement on the Great Plains. The images of sod houses and the determined people who built them were captured with the camera of Solomon D. Butcher. These photographs are familiar to every American who has seen a documentary film on the West, read a book about settling the Plains or just taken an American History class. For more than 50 years, the Butcher photographs have served as the illustration of home-

(Left) *Scan of the original glass-plate negative taken by Solomon D. Butcher of an unidentified family near West Union, Nebraska, 1886.*

(Top left) *Close-up of the door to the sod house in deep shadow. There is a hint of information, but not enough to make any real judgements on what is inside the soddie.*

(Top right) *Close-up of the door to the sod house, manipulated in Adobe PhotoShop to bring out the information in the shadow areas inside the house. Prior to scanning, the information inside the sod house was not available to researchers, as the prints they use show only a black doorway.*

(Far left) *Scan of the original glass-plate negative taken by Solomon D. Butcher of a grocery store in Overton, Nebraska, 1904.*

(Center) *Close-up of the doorway into the store, a very dark shadow area, with no recognizable detail.*

(Right) *The same close-up after some contrast manipulation in Adobe PhotoShop to bring out the information inside the shadows.*

steading on the Great Plains. We thought we knew everything there was to know about these photographs. We were wrong.

As we began scanning the plates, it immediately became apparent that there was more information in the photographs than initially realized. By the very nature of photography, negatives always hold more detail than prints. The negative is the first thing to be exposed, thus creating the photograph. The print is simply a mechanism to make the image more easily viewed. A black-and-white photographic print contains many shades of gray, from the deepest black in shadow areas to the whitest white in the highlights of the image. In a traditional darkroom, a printer must "pull-in" the highlights by printing them darker in order to maintain detail. The opposite is true for shadow areas. The printer must "hold" the shadows and make sure they do not get too dark, thus losing detail. A good black-and-white print shows black shadow areas with detail and white highlights with detail. This detail is only apparent, however, when captured in the negative.

Printing from 19th and early 20th century glass plates is especially difficult today because of the changes in manufacturing photographic papers. Glass-plate negatives have a tremendously broad tonal range. To illustrate the point, let's say that a glass plate has 250 different shades of gray. Today's photographic papers, however, may only allow one to create prints that have 175 or 200 shades. This immediately puts the printer at a disadvantage, making it practically impossible to accurately print a glass-plate negative. The results are prints that do not reveal all of the information that appears in the negative.

The advances made in the last several years in scanning technology have allowed for capturing increasingly higher resolution and density range. This translates into pulling more information from the original object during the scanning process than ever before. And more information in the electronic file leads to better understanding of the original.

In July 1998, when the Society began scanning the Butcher plates, we immediately realized the depth of information "hidden" in the plates that does not appear in the prints. On page 12 are some of the more poignant examples of that hidden information and how we revealed the rest of the history documented so well by Butcher over 100 years ago. The faces may be familiar, but the details are not. Remember as you look at these photographs, the information in those dark doorways has not been seen since Butcher released the shutter and created the photograph.

How We Did It

Creating an image file with this amount of detail requires scanning from the original in-camera negative. Dark shadow details on prints will not reveal hidden information if only the print is scanned. It also demands turning off any automatic settings on the scanning software.

Scanner Specifications

Hardware: Scitex EverSmart Pro™ Scanner
The EverSmart Pro™ is a high-resolution flatbed scanner that utilizes an 8000 pixel, anti-blooming CCD. It scans on both the X and Y axis enabling extremely high resolution over the entire image bed.

Resolution: up to 3,175 ppi optical resolution (uninterpolated)

Dynamic Range: 3.7

Tonal Depth: 42 bits per pixel [14 bits per channel, RGB]

Digital Conversion Criteria

File Format: Uncompressed TIFF files. A TIFF file is currently the most flexible file format for long-term storage of digital images. TIFFs can be turned into any other file formats required by our patrons, for example, JPGs or GIFs for World Wide Web viewing.

Resolution: The initial scans were 800 ppi uninterpolated, although resolutions as high as 3,175 ppi uninterpolated were used on several plates. These higher resolutions are most helpful for very small originals and images that are heavily used and most important intellectually.

Histogram/Clipping: Histograms offer the best view of tonal range captured by the scanner and are examined for clipping, as well as the correct setting of white and black points.

Clipping results when the white and black points are not set on true white and black during the set up of the scan. If white and black are improperly set, everything above or below those points is "clipped" or registers as the same tone. When clipping occurs, the scanning operator has misjudged the actual white or black points in the image and must rescan the original, using different settings for the white or black point.

Spiking on the ends of the histogram usually indicates clipping. This problem also shows up in the image itself as blockage and pixelization in the shadows and blowouts in the highlights. Acceptable spikes can occur if the edge of the original negative has lost emulsion, for example, or the sky holds no detail and is one tone in the original. Such instances, however, are rare.

Histogram/White & Black Points: Optimum placement of whites and blacks is best observed through the histogram. It is important to look at the number value assigned to the brightest highlight and the darkest shadow. Highlights should not read a number value higher than 247 and shadows should not be less than 7 or 8. If these numbers are exceeded, the scan must be redone. This is particularly vital if the original image has a short dynamic range, as is common with many Butcher negatives. The white and black points must not be set on zero and 255, as this will stretch the dynamic range of the image, creating gaps in the histograms, and thus unusable scans.

Mercuric Iodide Examples

(Above) *Scan of an original glass-plate negative, suffering from mercuric iodide intensification. This scan was created by telling the scanner the negative was actually a positive in order to show the appearance of the deterioration.*

(Above) *Scan of the same negative, but this time scanned as a negative and slightly adjusted in PhotoShop to increase the contrast of the original. Unidentified women in a buggy, c. 1890, from the photograph collections of the Nebraska State Historical Society.*

Scan of an original glass-plate negative, suffering from mercuric iodide intensification. This scan was created by telling the scanner the negative was actually a positive in order to show the appearance of the deterioration.

Scan of the original glass-plate negative, but this time scanned as a negative and slightly adjusted in PhotoShop to increase the contrast of the original. Photograph by George Condra of a group of unidentified men, possibly in western Nebraska, c. 1909.

Close-up of the Condra plate. Notice the Heinz Horseradish bottle in the lower right of the image. Prior to scanning, the information in this plate was irretrievable using traditional photographic methods.

The other important factors are resolution and dynamic range. The higher the resolution, the more small details held in the glass plate negative are recorded by the scanner. The larger the dynamic range, the more shades of gray can be detected by the scanner. Both elements contribute to revealing the hidden secrets of negatives.

Once a scan is complete, photo manipulation software is needed to change the contrast in selected areas of the print. This makes it possible to "pull-out" detail from the darkened areas on the image. Basically, the short tonal range that exists in a shadow area with detail is expanded to cover more tonal values, creating a higher contrast and making the information in the shadow area more easily seen. The Digital Imaging Laboratory uses Adobe PhotoShop to make these slight manipulations and reveal the hidden information in the shadow areas. It should be stated that the changes made to Master image files are not saved to the Master file. Any manipulation that is permanently saved is done so by creating a new image file.

Image Recovery from Deteriorating Glass-Plate Negatives

We discovered another benefit to digitizing that was not considered in the project planning. A common malady found in photographic collection, are negatives that have turned yellow due to mercuric iodide intensification. These negatives are very difficult to print in the darkroom, because the yellowed areas on the negative are virtually transparent. Modern photographic papers do not have the tonal range to detect the slight changes in value in the deteriorated yellow negatives. Even using grade 4 papers, which offer high contrast and straight Dectol, Kodak's photographic paper developer, normally used at 2:1, does not yield satisfactory results. Many institutions gave up on these plates years ago, forgetting about them or in some cases disposing of them completely. Digital imaging turns these once unusable negatives into viable research tools. For more information about image recovery from deteriorating glass-plate negatives, go to the Nebraska State Historical Society's Web page: *http://www.nebraskahistory.org/lib-arch/research/photos/digital/preserve.htm*

As we began scanning the Butcher negatives for the American Memory Project, we found one negative that was suffering from mercuric iodide intensification. We decided to try and scan the yellowed plate as a test. In the past, attempts to print this plate in a darkroom were not successful, so we thought scanning the plate might work. We were very pleased with the results. The scanner was able to record the subtle differences between tonal values in the affected areas of the image enabling information retrieval.

Since that first test, we have scanned numerous negatives from the collections that are yellowed from mercuric iodide intensification. No manipulation was necessary for the scanner to detect the shades of yellow. We scanned the plates as if they were unaffected black-and-white negatives, and the scanner looked at the shifts in tonal value and recorded a positive grayscale image. In some instances, the resulting Master scan was slightly flat, however, we found that to be the case on many of the Butcher images. Slight adjustments to contrast in PhotoShop allowed us to produce high quality prints and surrogate image files for posting on the Web. The Master files remain slightly flat, as our goal is to produce a Master scan that is as close a representation of the original as possible, knowing we can always make adjustments to the file for any use in the future. If we adjust the Master, we lose our zero point from which to start. Prior to scanning, these yellow negatives were virtually worthless. Now they are an important part of the collection.

Why This Is Important

Aside from the obvious benefits of allowing access to every Butcher image on a Web site, and the fact that digital image files are so much better than the microfiche technology that hundreds of researchers have borrowed from the Society since the mid-1980s, we finally have the opportunity to offer access to information that was available, but practically impossible to use. No one ever thought about looking at the negatives, because the prints were available. And those prints are loaded with good information about homesteading. The question never came up to suggest that there might be more to the images than met the eye. Until, that is, we started scanning.

The implications on historical research are mind-boggling. The ability to look inside a sod house was very limited, until we scanned this collection. Very few photographs were taken inside a soddie, so opening these dark doorways and looking inside changes everything for the researcher. Material culture studies will allow for statistical examination of the interior of sod houses. Beds are in front of open doors to enable more comfortable sleeping in the hot Nebraska summers. Made beds are found more frequently where a family is photographed by the soddie, rather than a bachelor. Small details barely apparent and usually overlooked are now easily highlighted and recognized. There are many more people in these images than we ever supposed. Children are peeking around the corner of the soddies. Other children are inside the house, while a family is posed outside for the photograph. Who are these children? Are they neighbors or children of the hired help? Speculation and educated guessing is rampant regarding these intimate details revealed for the first time. Why are people peeking around buildings, skulking in doorways and windows, and hiding behind the livestock?

Digital technology is opening new doors for photographic and material culture research. It is time to take another look at those yellow negatives in your collections, at historical research and most importantly at the significance of scanning the in-camera negative.

For more information about the Digital Imaging Laboratory at the Nebraska State Historical Society, visit http://www.nebraskahistory.org/lib-arch/research/photos/digital/index.htm

Portions of this article previously appeared in the Summer Issue 2000 of Nebraska History *magazine.*

Jill Marie Koelling
Curator of Photographs
Head, Digital Imaging
Nebraska State Historical Society

Jill Marie Koelling (koelling@nebraskahistory.org) is Curator of Photographs and Head of Digital Imaging at the Nebraska State Historical Society. She has a BA in Photography from Montana State University and an MA in Museum Studies from the University of Nebraska. Koelling's research interests include the development of early tourism in the United States and the impact of snapshot photography on American culture.

"Building a Digital Archive – Digitizing for the Long Term"

Carol Hernandez and Robin Lilien
The J. Paul Getty Museum

Creating a digital archive is quickly becoming an integral part of museum management. How are digital images used in the museum environment? What are the benefits versus the cost and time involved in creating an archive? How do you build an archive that will be useful for the long term?

At first glance, the task appears overwhelming. But, as in any major project, a successful strategy involves careful planning and a thorough analysis of needs and expectations.

The uses for a digital archive in a museum environment are extensive. Digital images are beneficial across a wide range of museum activities, such as curatorial, education and conservation. For curatorial staff, digital images can be used for preservation, identification, research and collections management. They are also helpful for exhibition planning and installation. For education, greater exposure of the collection is possible as well as online study by students and scholars. In conservation, digital images are useful for documentation and study via electronic image manipulation.

Among the many benefits derived from a digital archive is reduced object handling, which decreases deterioration of the collection and, consequently, preserves the object. Research can be performed without looking for and moving objects in storage as the archive provides access to the object's surrogate images.

Where do you begin? Start with the big picture if you want to address a long-term solution. Seek a comprehensive solution as opposed to a series of one-off projects. Here is an example of what happens when you do scans for one-off projects. You can scan the image for use on the Web and save it in JPEG format. Then you could have another project to create a CD, so you scan it at a little higher resolution and save it in BMP format. Finally, you scan the same image for exhibition materials at a much higher resolution and you save it in TIFF format. If you had done the higher-resolution TIFF image to start with, there would be no need to rescan each time. You would have saved time overall, and you would have consistency across all images.

Create a comprehensive needs statement to determine the scope and purpose of the digital archive. Survey all digital-imaging projects proposed across the museum and discuss specific needs and expectations. Decide how thoroughly the collection will be represented and how many images will be digitized. Knowing your needs will make the rest of the process much easier, because you will know the amount of time needed to complete the project and the level of image quality you require. Most public and educational uses require medium-quality images, while preservation, research and repurposing require high-quality images. Keep in mind the original mission of the museum as you do this. A request for substantial funding, which most large-scale digital imaging projects will require, must tie back to the core functions of your institution.

Your quality requirements will enable you to determine the technical specifications for the archive. For a high-quality archive where you will achieve the maximum cost benefit over an extended period of time, it is recommended that the images be scanned at a high resolution. For example: a 4x5 transparency scanned at 2,000 ppi. Keep in mind that the smaller the original, the higher the ppi for your initial scan. This archival quality scan should be stored off-line, then sampled down to a maximum of 40 MB files for any online use. For instance, a single high-quality scan of a 4x5 transparency will yield a 200 MB file for preservation, a 100 MB file for high-end publishing, a 40 MB file for an educational kiosk and various smaller files for other online use.

In a high-quality archive, the resolution of the digital image should be as good as the source transparency. The initial image should be scanned "raw" using a well-calibrated scanner. The image should be scanned as accurately as possible and, at this stage, should not be taken into additional image editing software for color correction. It should contain maximum detail in both the highlight and shadow areas and should be saved as an uncompressed TIFF file.

Your comprehensive needs statement will give you vital information to proceed to the next important step: creating a budget and performing an accurate cost analysis.

You will now be able to determine the equipment needed to meet technical specifications, set production goals and determine priorities as to which objects will be digitized first.

Decide how you will track the images once you start building your archive. If it is at all possible, have the Collections System track the metadata about the digital images: file characteristics such as name, location, type, size, color, format, creation date, etc. If it is a derivative, the source file from which it was created. If the archive is being created from photographic media, it will be necessary to create metadata for this media first. After that, you can link your digital file information to this data. If your Collections Management System cannot accommodate this new data, there is a variety of digital asset management software available for this purpose.

Obtain estimates from outside vendors. Determine the internal staffing needed to coordinate working with a vendor. At the same time, analyze the staff requirements, equipment specifications and room specifications to create the archive in-house. With this information, you can determine whether it would be more cost-effective for your museum to outsource or create an in-house lab, or if necessary, divide the work between the two. After an extensive vendor evaluation, our institution, given our long-term requirements, found an in-house lab would provide greater quality and technical control. It has significantly saved money and time.

Even if you select to outsource, there are options for how to divide the work between you and your vendor. You may decide to outsource all of the technical work, including archiving to tape or optical storage, creating smaller file versions, correcting these derivative images, and outputting to CD or other media. Or, you may elect to handle some or all of this post-scanning work in-house.

If you decide to set up your own in-house lab, buy the best quality equipment you can afford. Choose a method for capturing the original material: direct capture or film scanning. Both methods have positive and negative aspects. There are many issues to be considered in making this decision. How fragile is the original material? How much of your existing traditional photographic equipment can be incorporated if direct capture is utilized? How many high-quality original transparencies of the collection are available? Direct capture is a relatively new and untested method for creating a high-end archive. There are new issues that arise. Staff photographers will need to do some research to gain expertise in the nuances of direct capture. The CCD (Charged-Couple Device) of the digital camera sees objects differently than traditional film. Depth of field is an issue when shooting three-dimensional objects. The effect of long exposure of the objects to the lighting required for direct capture is unknown. The two main advantages of direct capture are its immediacy and the reduction in film and processing costs. Test both methods thoroughly using your final output requirements as a guide. This testing will be key in determining which methods will be suitable for your purposes and you can also determine how to divide the work between the two methods, if appropriate.

When purchasing a scanner, it is important to consider the dynamic range of the machine. The dynamic range is the ability of the device to record detail in both the shadow and highlight areas of an image. It is also important to know the optical resolution of the device. Make sure the manufacturer has specified how much of its resolution claim is interpolated (artificial pixels adding "resolution" via software algorithms). Also consider the bit depth, maximum scanning area and quality of the software interface which operates the scanner. There are new flatbed scanners on the market that are starting to yield quality scans as good as a traditional drum scanner. Test equipment thoroughly before purchasing it using representative images in the collections. Most hardware vendors are very willing to accommodate this for no charge. Based on test performance, compute the anticipated through-put and be sure to include preparation time and not just scan time. Put these test images through the various output scenarios you require: on screen, to print and to film. The importance of this testing cannot be overemphasized. Know that the equipment will give you the performance you require before you buy it.

Budget, priorities, quality requirements and resulting technical specifications will determine staffing. Evolve your digital team and its management. The critical element in the control of quality is hiring experienced professionals in the key positions of manager and senior technician. These individuals will be required to have prior experience with image-editing software, color management and scanning. Key staff should be familiar with the characteristics of digital files required for a wide variety of output. Most importantly, they should be familiar with the digital requirements you have determined through the museum 's comprehensive needs statement. You will also need to have access to a technically oriented clerical assistant to deal with the tracking and trafficking of the workload. New staff can be added over time as required, but keep in mind that each addition or modification to your process impacts the entire workflow and requires adjustment.

This is an exciting time and a time where new opportunities and avenues of access are opening up for museums and the community they serve. With careful planning and a thorough analysis of needs and expectations, your digital archive will provide use for many, many years to come.

Carol Hernandez
Manager of Digital Imaging
The J. Paul Getty Museum
Carol Hernandez (chernandez@getty.edu) is the Manager of Digital Imaging at the J. Paul Getty Museum in Los Angeles, California. She has a degree in photography from Rochester Institute of Technology.

Robin Lilien
Manager of Information and Media Systems
The J. Paul Getty Museum
Robin Lilien (RLilien@getty.edu) is the Manager of Information and Media Systems at the J. Paul Getty Museum. Prior to coming to the Getty 10 years ago, she was a manager in systems development at Arthur Andersen in Chicago.

What are Digital Image Managers?

Dr. S. K. Hastings
School of Library and Information Sciences, University of North Texas

Introduction

The University of North Texas in Denton, the Dallas Museum of Art, the African American Museum in Dallas and the Denton Public Library System are partners in a program designed to produce a cadre of experts in the field of digital image management. With funding from the federal Institute of Museum and Library Services, the School of Library and Information Sciences, has built up a digital imaging laboratory, houses and maintains the African American Museum Web site, and provides fellowships for students in three degree programs. Each of the degree programs—Masters in Library and Information Sciences, Certificate of Advanced Study and Interdisciplinary Ph.D. in Information Science—offer a specialty in Digital Image Management. The School of Visual Arts adds experience in museum education, use of computers in art, graphic design and aids in the digitization of images throughout the programs.

The project includes the production of digital images, digital information database creation and management with a focus on advanced network and information technologies. The production and management of digital images and the management of digital information are important skill-sets for current and future library and museum information professionals. In addition, the program prepares individuals to assume positions as experts in the broader markets of libraries, archives and information centers. An integral part of the educational experience for the students enrolled in the program is the opportunity to work as interns at the museums and libraries. One result of the project is improved access to our cultural heritage in the networked environment.

My year as an IMLS fellow has taught me a great many things. I always tell people that the most important lessons I learned were in group dynamics. This very diverse group of talents, personalities and backgrounds forced me to compromise, to be tolerant and to understand that every opinion is worth hearing.

My favourite part of the digital image lab was going through the Sepia collection. Scanning and data entry became less tedious and more educational with every image I scanned that wasn't part of a "Fashions" folder. I especially appreciated working with software and hardware that was previously foreign to me, like Dreamweaver and the digital camera. I even treasure the experiences I did not enjoy, like finalizing the design of a Web site in three days with new software and 10 people on two workstations.

The opportunities to attend conferences and travel will not soon be forgotten.

Jodi Kearns, Canada

Why Digital Image Managers?

Information has become a critical resource in most facets of American society, and education for the information professions has become strategic in the development of a healthy information economy and the preservation of democracy. Digital image and information management is a subset of this much larger construct. In order to compete and lead in a global economic environment, we must have experts capable of planning, implementing and managing projects that increase and improve access to our cultural heritage. This project focuses on the importance of digital information in a variety of organizations dedicated to preserving and presenting cultural objects.

Specifically, we have an obligation to protect and preserve the cultural heritage that we have spent billions of dollars collecting. It is not enough to have staff capable of creating digital surrogates of our collections. We must invest in personnel that know how to manage and preserve these assets.

In an e-mail exchange discussing a draft of the present article, *Spectra* Guest Editor Günter Waibel suggested that Digital Image Managers could be thought of as the "registrars" of the digital realm. While I think this is true, I also do believe that we need to think beyond the walls of the institutions that have been the traditional guardians of our cultural objects. What is the equivalent of a registrar in our academies, in our libraries, in our information-based businesses? I believe it is the leader, the visionary, the technologist, the researcher with the unusual combination of skill-sets that both produces and manages our digital assets.

What are Digital Image Managers?

Günter also asked me what I thought to be vital parts of the educa-

tion for a Digital Image Manager or a Manager of Digital Media or a Digital Assets Manager or a Visual Resources Manager or a Digital Archivist or _____. Well, you get the picture. We have many names for this new breed of librarians, resource managers and archivists. It is the digital environment and the importance of our approach to 'all things digital' that really define what Digital Image Managers do. Rather than go to the academic catalog of courses and curricula, I thought you might like to hear from a few of our fellows what this educational opportunity has meant to them. Please refer to the sidebars throughout the article for unique views about the program from the IMLS fellows.

You will see from the educational experiences depicted by the Fellows that we are trying to produce people with an eclectic set of tools – from being able to lead a diverse team to understanding the philosophical issues of representation with digital images. A list of technological competencies bridges the leadership responsibilities we stress. The competencies fall into four categories, although they are not always distinct.

We start with the selection, creation and acquisition of images including photography, lighting, digital cameras, scanners and image formats. Included here are the discussions of collection development, intellectual property and copyright issues. Where does the digital collection fit in the mission of the institution? What are the costs and benefits of digital signatures, watermarks and software that prevents images from being copied?

Next, we look at the effects of manipulation on the images. What happens if you keep cropping an image? How much information is lost? What is the relationship between image format and resolution? How do we design for both high-end and low-end display units? What steps are involved in processing the digital image, from thumbnails to archival formats? We look at how to store and maintain the archival image record of the object, the cost of doing so and the importance of migration and preservation planning.

Third, we investigate how to organize and retrieve the images. What

When I was granted a fellowship from the Institute of Museum and Library Services for the semesters of Fall 1999 and Spring/Summer 2000, little did I know how pivotal those semesters would be for me in terms of my professional and my personal life. The fellowship granted me entrance into Dr. Sam Hastings Digital Imaging class, the end result to be a Certificate of Advanced Study in the area of Digital Image Management.

Professionally, the experience prepared me to manage the planning and establishment of any digital imaging project. In terms of preparation, the class introduced me to the questions that must be answered and the various paradigms that might be considered when approaching the different aspects of the planning process. I was also able to build a list of information resources from which to draw in answering those questions under any variety of project conditions.

Though I am still (and always will be) learning about the various technologies available for implementing such a project, both in terms of hardware and software, the class also introduced me to the resources I can call upon to garner the information I need concerning the technical possibilities available to me on any given project. It also prepared me in a fundamental way to utilize those technical resources in such a way that I can insure a successful outcome.

Personally, the digital imaging class provided growth for me in terms of working with others in teams. It was an excellent opportunity to learn various ways to meet the challenge of determining the best solution (through teamwork) from a large set of very good possible solutions. It was also indeed a pleasure to work with other people from all over the world who each provided their own special brand of professional talent and human kindness.

Marsha Merritt, Texas USA

As a summary, here's a general job description for a Digital Image Manager.

TITLE: Digital Image Manager

RESPONSIBILITIES: "Digital Assets Incorporated" or "Next Generation Museum" or "Crystal Light Library" is seeking a knowledgeable, creative and highly motivated individual who can contribute to the development, application and maintenance of our digital image resources. This position will provide leadership in exploring image-related digital library initiatives and implementation strategies. The position will serve as liaison to the Research and Development Department, where activities may include product development and image retrieval research. Collaboration with company information specialists in a team environment to introduce or enhance digital access to image resources is an important facet of this position. Additional responsibilities include: selection and acquisition of images, liaison to the legal department, creation and maintenance of digital image databases, initiatives for preservation and access, designing modules for training of all staff, design and maintenance of Web access, as well as liaison to the Information Technology Department.

QUALIFICATIONS: MLS degree with specialty in digital image management from an ALA-accredited institution is required, art history background or degree in museum studies is preferred. Applicants must have a strong interest in leading research and development efforts related to digital image retrieval, user interface design and the integration of databases and images in a digital library environment. Candidates for this position must also have demonstrated skills in the uses and applications of database systems, image manipulation software, indexing and metadata, search engine design and development, and popular operating systems such as NT or UNIX. Must also possess excellent communication skills, both written and oral, and the ability to work collegially in a multicultural team environment.

SALARY: Minimum starting salary $45,000, negotiable depending on qualifications.

types of metadata are needed? How well do new mark-up languages, such as XML, work with images? What depth of indexing do we apply? What database fields should be included in what type of database? How do search engines work with relational databases? How do we build Web-based search engines? How effective are content-based search engines in non-homogenous collections? Issues of vocabulary and control are investigated and the use of traditional thesauri for contemporary works is discussed. We look at the problems inherent in image retrieval and focus on the need for browsing large blocks of images, user relevance feedback and methods for user need analysis. In addition, we try to step back and address the issues of quality control, periodic evaluation and revision, and project workflow.

Finally, we design, build and maintain Web access to the images and databases created. Our partnership with the School of Visual Arts (SOVA) is invaluable to this category of work. Graphic design and the "artistic" eye are not always inherent skill-sets in library and information science professionals. We rely heavily on the fellows from SOVA to lead us through the design phases for the Web sites. Included here are issues of networking and data communications, as well as staff training at our partner institutions. Project evaluation, as well as user evaluation, is part of the access issues category. How do we know when we are successful? How do we collect data that tells us that we have made a difference in someone's life? The IMLS projects are leaders in the use of Performance Evaluation. If you are interested in finding out more, please refer to the resources at *http://www.imls.gov*. We also discuss public relations and marketing, and work with the museum staff to design and implement publicity packages.

Where Do Digital Image Managers Work?

So, what does all this mean in today's job market? In addition to the traditional employment opportunities represented by museums, libraries, archives and other information-based institutions, we are finding that the corporate sector is extremely interested in Digital Image Managers. Companies are quickly beginning to realize that the meticulous management of their digital assets ensures a greater financial return. Information Architecture firms were some of the first to hire our graduates, and they continue to request graduates with experience in high-end digital resources. In addition, high-tech industries with digital initiatives have become regular raiders of our workforce. Several library directors have lamented the loss of their information technology personnel to the corporate sector. Photo stock companies love our graduates and often call to see if we have more ready to hire. Comments from employers focus on how valuable the combination of leadership with technology skill-sets has become in product development, as well as management and preservation of existing assets.

Naturally, I want to see our graduates placed in successful employment. As altruistic as most of us are, our students will not continue to take jobs in the nonprofit sector just because it is the "right thing to do." If we want to hire this level of expertise, we need to be somewhat competitive in our salary ranges. When libraries and museums talk about return on investment, there is a whole lot more at stake than just money. It is difficult to put a price on cultural heritage and knowledge. It is up to us to ensure that the digital assets we so painstakingly build will be available for our children and their children.

My participation in the Digital Image Management Project with diversified fellow students has been a golden opportunity for me! The program provided me with specialization in all aspects of the management of digital images, including production of digital images, database creation and management and maintaining Web access for a digital collection.

I believe that these skills have enhanced my abilities and provided me with more opportunities in the current job market. Currently, I am working in the government documents department at the University of North Texas Library on the Digitization and Documents Electronic Access Program. The project is affectionately called the "Digital Cemetery."

I strongly believe that the digital image management skills I gained from the program coupled with my professional experience in library and information management areas help me understand, develop, expand on and use opportunities inherent in libraries and museums by creating digital images of library resources and museum objects.

Finally, I would like to take this opportunity to thank the Institute of Museum and Library Services for the continued support which made my participation in the Program possible. I am looking forward to working on the Interdisciplinary Ph.D. in Information Science, with a particular emphasis in Digital Image Management.

Daniel Gelaw Alemneh, Ethiopia

What Are We Doing at UNT/SLIS?

The University of North Texas, School of Library and Information Sciences program offers a highly effective structured curriculum. The energies of the faculty are focused on information-related problems, including the retrieval of digital images and the management and organization of digital information. Thanks to the support and funding from IMLS, we are preparing program information, publicity and information packages and will announce the Digital Image Managers programs of study and fellowship opportunities in late fall 2000. What follows is a progress report on the main areas of activity in this ongoing project.

Criteria and Process for Selection of Fellows

Applicants for the Certificate of Advanced Study (CAS) met the general admission requirements for the school which include: holding a master's degree in Library and Information Sciences from an ALA-accredited school with a grade point average of 3.0 or above, and an advisor's recommendation for program of study. In addition, the candidates demonstrate a strong motivation toward pursuit of a career in digital image management through work experience, personal essays and interviews. Candidates promise to make significant contributions to the field and to complete the CAS as evidenced by career progression, prior research and publication and/or professional activities. The fellowships were awarded to candidates that showed a need for financial support in this endeavor, as illustrated by documented projected expenses and assessment of available personal resources as determined by standardized measures specified by the granting agency. In spring 1999, we recruited 10 fellows for the program of study to begin in fall 1999. Approximately half of the fellowships went to international students.

Taking a variety of classes provided by SLIS and SOVA enabled me to understand standards, protocols for information retrieval (Z39.50), economics of information, digital imaging, management of digital collections, HTML, XML, issues of educational programs as well as collection preservation in museums.

I have learned more in the 1999-2000 academic year than I have learned during my entire life. I enhanced my skills with Web development tools and techniques (client-side scripting, Web site maintenance), database management tools (Microsoft Access) working with different computer environments (UNIX, Windows NT, Windows 95/98). I learned from scratch such programs as Adobe Photoshop 5.5, Adobe ImageReady, Macromedia Dreamweaver 3.3, Fireworks and Frontpage 2000. It was a privilege to have our own digital lab where I could spend as much time as I wanted scanning photographs, entering data into the database, and experimenting with Adobe Photoshop 5.5.

I was able to see many of the great museum collections in the Dallas Fort Worth area. Besides visiting the current exhibits, I was allowed to visit storage rooms and the preservation rooms, which are not accessible to the regular museum visitor.

Working with 10 fellows from six countries taught me how to work better as a team, with a lesson in patience, and helped me to learn the art of compromise. Working in an international environment enriched my knowledge of other cultures and taught me skills to help me in my future work experiences.

Victoria Kravchyna, Ukraine

In fall 2000, we will begin recruiting for the masters and Ph.D. programs, as well as add further CAS candidates.

Museum Data and Impact Study Design

One of the most interesting problems presented in the project revolves around measuring the impact that a Web site of digital images may have on the museum. Questions of how to capture baseline museum data so we can look at possible impact are very challenging. As we assess the museum collections to identify and select the first batch of objects to be digitized, we will be building a description of the collections and a history of when and where each object was shown. We know that approximately 80 percent of the collections have not been shown and this seems to be a standard percentage for museums. When all of the museum's collection is available for identification on the Web site, will it increase the number of requests that the museum receives to view an object? What impact would the increased number of requests have on museum operations? How many and what types of visitors does the museum currently have? How many and what types of visitors will the Web site have? These are questions that require data collection instruments that can be used across the board. Current output measures that we are collecting include the number and kinds of programs and services distributed by the museum and the audience served.

A large part of the museum's mission is directed toward educational programs that include summer camps, school tours and exhibits. The design of the Web site will include educational activities and curricula built around the digital image collections and information database. What types of comparisons will we be able to draw from these similar, but unique experiences? Part of the educational experience for the fellows in the program is the opportunity to conduct research, to design effective evaluation methods for the museum, to collect and analyze data and to write research reports. Mr. Miguel Arroyo Morales is currently looking at the problems and opportunities of collaboration as his dissertation research.

Digital Image Laboratory Equipment

Networked workstations with flatbed and slide scanners, a CD-ROM burner and a color laser printer serve as the main equipment in the digital image laboratory. A server with 40 gigabits of storage is running Windows NT and acts as the primary Web server for the project. The lab is connected to the museums by a V-Tel video-conference system and ISDN lines. With the new grant funds, we will be adding 12 new workstations and peripherals. The lab is administrated and maintained by Schools of Library and Information Sciences and Visual Arts and the Academic Computing Services. The lab manager has prepared policies, instructions and guidelines for the use of equipment in the digital image laboratory and designed a checklist for students to use in the lab. Access to the lab is currently restricted to students in the digital image management program of study, but we hope to open the lab for other classes eventually.

Evaluation

Measures of success of the programs of study include:

- Statistics on the number of inquiries received and number of applications submitted versus the number of enrollees;
- Profiles of all fellows including academic performance and placement information;
- Career progress and accomplishments including publications, participation in professional associations, etc.

The overall plan for ongoing evaluation has three parts.

- Primary evaluation for the project is based on meeting the measurable objectives outlined in the goals and objectives of the project plan. Primary evaluation includes measures of success of the program graduates and provides opportunities to study the impact of well-prepared professionals in the field.
- Secondary evaluation collects data to produce models of the program of study and the cooperation and collaboration between libraries, museums and universities.
- The third level of evaluation is ongoing and tracks the impact of increasing access to a museum collection by providing digital images on the Web.

Summary

This paper describes what makes a Digital Image Manager and the importance they may have in the field. Several of the IMLS fellows have provided their views of their educational milestones while working with the wonderful collections from the African American Museum. Highlights include learning image manipulation software, image acquisition, Web site development, database construction, representation and metadata, and strong beliefs in the power of teamwork.

As an educational experience, one of the most unique features of the project is made possible by the cooperation between libraries, museums and a university. Students digitize objects from the museums' collections and build image databases available on the Web. The experience includes the use of video-conference technologies to enhance collaboration between the partners. Students communicate with the museum staff, work on images and database information simultaneously, and share experiences immediately.

The project is a model for training professionals in the creation, use and management of digital images and may serve as a model of collaboration between museums and universities. In addition, the project assesses the impact resulting from improved access to museum collections through the Web.

The project is ongoing and scheduled to conclude in fall 2002. Preliminary data reports and in-progress evaluations are available at the project Web site, linked from *http://www.unt.edu/slis*. The African American Museum Web site we built is at *http://www.aamdallas.org*. Comments and suggestions are encouraged and should be addressed to the author. Special appreciation goes to the federal Institute of Museum and Library Services and the University of North Texas for the funds to make this project possible.

Dr. Samantha K. Hastings
University of North Texas
School of Library and Information Sciences

Dr. S. K. Hastings (hastings@lis.unt.edu) joined the faculty at the University of North Texas in Denton in 1995. Sam's research interests are the retrieval of digital images, telecommunications and evaluation of networked information services. Sam tries to integrate real-world experiences as reflected by teamwork and product development in all of her classes, which range from indexing and abstracting to telecommunications. Prior to moving to Texas, Sam worked for the State Library of Florida as the network and telecommunications consultant for statewide, multi-type library networks. She received her doctorate in Library and Information Science from Florida State University in 1994. She has over 20 years of experience in libraries and information management. Sam has worked as a consultant and built full-text and image databases for accountants, dentists, doctors, lawyers, and county and state governments.

Take advantage of the power and presence of a display ad in the MCN Spectra

Get your message out to more than 700 MCN Members and 600 Museums worldwide by advertising in the MCN Spectra.

	1x	2x	3x	4x
Full Page (8” x 10 5/8”)	$300	$300/each	$200/each	$200/each
Half Page (8” x 5 1/4”)	$150	$150/each	$100/each	$100/each
Special Positioning				
Inside Front Cover	$400	$400/each	$250/each	$250/each
Inside Back Cover	$400	$400/each	$250/each	$250/each
Outside Back Cover	$600	$600/each	$400/each	$400/each

PAYMENT: MCN requires prepayment of all ads at the time of placement. Please submit your payment along with the insertion order on or before the appropriate deadline.

DEADLINES: The MCN Spectra is published four times a year in Spring, Summer, Fall and Winter.Your Insertion Order and Payment are required to make your confirmed space reservation. Your Artwork is due one month after space deadline. The 2001 schedule is as follows:

Winter Issue (mails mid-January) — November 1: Spring Issue (mails mid-April) — February 1

Summer Issue (mails mid-August) — June 1: Fall Issue (mails mid-November) — September 1

MECHANCIAL REQUIREMENTS: Black-and-white electronic artwork is preferred. Film is also accepted when supplied right-reading, emulsion side down. Please set up electronic files in Macinstosh Quark Xpress or as an EPS file from Adobe Illsutrator or Adobe Photoshop.

PREPARATION ASSISTANCE: A $60/hour fee will be charged for art and composition services if needed.

QUESTIONS: Please contact Kathleen Buckner, production manager, at 877/626-3800, ext. 210 or via e-mail at kbuckner@mcn.edu.

INSERTION ORDER FORM CAN BE FAXED TO YOU BY CONTACTING KATHLEEN BUCKNER AT 877/626-3800, EXT. 210 OR VIA E-MAIL AT KBUCKNER@MCN.EDU.

Light Exposure to Sensitive Artworks During Digital Photography

By Ben Blackwell

As more museums consider plans to digitize their collections, there have been concerns about the light levels required by digital scanning cameras. There has been little published on the subject, or indeed on exposure during conventional photography. Most museums, however, follow guidelines governing light exposure during exhibitions. Since exposure follows the reciprocity law, it's easy to translate exposure during photography into its equivalent under gallery conditions.

Reciprocity

Bunsen and Roscoe's law, expressed as H = l x t = constant, implies that exposure, (H), at a high intensity for a short time (t), is the same as exposure at a low intensity (l) for a correspondingly longer time (t). 1000 lux for 10 seconds has the same effect as 10 lux for 1000 seconds. Or, as expressed by Stefan Michalski, "Light effects are cumulative in a simple additive manner."[1] According to Michalski, a number of studies have upheld this principle as applied to light-sensitive materials, notably that conducted by Saunders and Kirby of the National Gallery of London.[2]

Most photographers are familiar with the reciprocity law, on which equivalent exposures are based, and with the notion of reciprocity failure— the fact that some photosensitive materials don't respond as predicted at very short or very long exposures. These deviations are always in the direction of less effect, and aren't relevant to this discussion.

Museum Policies for Exhibition Exposure

Most museums monitor the environment in which their collections are kept, and impose restrictions on the type and amount of light in the galleries, and on the length of exhibitions. These rules vary, but for works on paper, photographs and other sensitive objects, five-foot-candles or 50 lux is a common standard for gallery lighting.[3] This is quite dim for reading, working or critical seeing.[4] Guidelines for exhibition length range from four to 12 weeks per year, six weeks being average.

In 1991, Karen Colby developed an exhibition policy for the Montreal Museum of Art, which has been widely cited.[5] Based on the British Blue Wool or International Standards Organization (ISO)[6] standard for light-induced fading, she divides works into three categories of sensitivity and assigns exhibition light levels and durations accordingly. For Category 1 objects, the most fugitive, she recommends a maximum of four weeks exhibition per year at no more than 75 lux. This amounts to 12,000 lux/hours per year based on the 42-hour exhibition week at Montreal. The 75 lux standard is more liberal than some, but allows for some flexibility in exhibition design and the widespread skepticism about other institutions actually adhering to 50 lux for traveling exhibitions. At this level of exposure, works in this category would show just noticeable fading after 1.2 Mlux/hours, or 100 years of annual four-week exhibitions.[7] Works on paper assigned to categories 2 and 3 are allowed 100 lux for 10 and 20 weeks respectively, and fading would be noticed at 250 years (10 Mlux/hours) and 3500 years (300 Mlux hours).

It's clear that works on paper, considered among the most fragile and fugitive artworks, vary greatly in their vulnerability to light. However, Colby's Category 1 is a useful benchmark for assessing light exposure during photography. Category 1 includes materials that fall into the ISO levels 1, 2 and 3, but is based on the middle of that range. ISO 1 includes some objects that can tolerate virtually no light exposure at all, and these must be assessed individually as to whether they may be photographed or even viewed. Works this sensitive that haven't already faded to nothing represent a small percentage in most museum collections.

Types of Digital Cameras

Institutions considering direct digital capture of artworks have several options. Small works on paper can be scanned in flatbed scanners[8], but for a wider range of objects, some sort of digital camera is required. Digital cameras come in several basic designs and a wide range of prices, any of which might find some application in museums. Institutions creating high-resolution digital images adequate for various purposes, including fine-print publication, are mostly using scanning cameras. A smaller number use high-end area-array cameras.

Area-array, or "chip" cameras, ranging from the familiar consumer digital cameras to professional models costing $30,000 or more, record images on a square or rectangular array of electronic sensors, usually Charged-Couple Devices (CCDs). Models capable of "instantaneous" capture are able to photograph moving subjects, but are limited in resolution to a maximum of about 2000 X 3000 pixels. Moreover, since each sensor is dedicated to recording a single primary color of light— red, green or blue—the other two colors for each pixel have to be interpolated from data supplied by the adjacent sensors. "Three-shot" cameras avoid interpolation by taking separate exposures through red, green and blue filters. These can capture monochrome images of moving subjects, but color images of stationary objects only. Maximum image size is the same as for high-end one-shot cameras —2000 X 3000 pixels—producing an 18 MB file. Both types can be used with electronic flash, an ad-

vantage when cumulative exposure is a concern. Some institutions are using three-shot cameras, but because they're expensive and produce a relatively small file, they're not as common as scanning cameras in museums and archives.[9]

Trilinear scanning cameras, or scan backs, incorporate three rows of sensors that travel across the image area. Because a relatively small number of CCDs record data across a large area, they produce the highest resolution images of any digital camera. The top-end models cost $25,000 to $30,000, about the same as the high-end chip cameras, but produce files up to around 500 MB. Mid-range models at a little more than half the price still render uninterpolated images of around 6000 X 8000 pixels, producing 24 bit RGB files of 140 MB. Since they work like a scanner inserted into a view camera in place of the film holder (or a medium-format film back in some smaller models), they require continuous light. Light requirements are higher than for shooting film under continuous light because during even a long scan, each data point is only exposed for a brief time (usually 1/8 second maximum), whereas during a long film exposure, the entire emulsion accumulates exposure at once. Exposure is greater than in a flatbed scanner, because the lights must be on for focusing and other adjustments, as well as for the scan. Some of these operations, however, may be carried out under reduced light.

Light Requirements of Digital Scan Backs

To determine the amount of exposure required by digital scan backs, I've conducted tests at the UC Berkeley Art Museum & Pacific Film Archive using a Better Light Model Super 6K. This mid-range model is capable of capturing images of 6000 X 8000 pixels at standard resolution, or 9000 X 12000 pixels with minimal interpolation across the shorter dimension. Images can also be recorded at a number of lower resolutions, and this has a bearing on light exposure as scans become correspondingly shorter at lower resolutions. Scans can take from under a minute to 15 minutes or more, depending on light intensity, resolution and other camera settings. Time under the lights for framing, focusing, prescans and adjustments vary even more, depending on the nature of the subject and setup and the operator's methods and skill. For flat copy work, we can assume between 10 and 20 minutes under the lights, based on the variables mentioned.[10] Beside my own experience, observing other operators using similar equipment suggests 20 minutes as a reasonable maximum for flat copy work. For production copying of similar size originals, times can be much shorter, and for complicated three-dimensional subjects, much longer.

Line time (analogous to shutter speed) and ISO settings (equivalent to film speed) are adjustable. Increasing either compensates for less light. However, increasing line time results in longer scan time and increased noise. Higher ISO settings increase noise at an even faster rate.[11] Most digital photographers prefer to work at the lowest ISO and the shortest line time light levels permit for shorter scans and cleaner images.

Visiting other studios, I've generally encountered higher light levels than the circa 2000 lux I've been working with. With more light, scans are shortened, but other operations carried out under the lights

The following tables show the exposure in lux/hours for captures at different resolutions for two actual copy setups, one with florescent light and one with tungsten.

Copy set up for 16x20 original, 2 Balcar florescent fixtures @ 6 feet. Illumination at copy stage 2056 lux (190 foot-candles) Camera settings: aperture f11, ISO 400, line time 1/20 sec.

Pixel dimensions	File size	Scan time in minutes	Total time under lights in minutes	Exposure in lux/hours	Equivalent gallery time at 50 lux
12000 x 9000	309 MB	10	20	678	13 hours, 42 minutes
8000 x 6000	137.3 MB	6:40	16:40	570	11 hours, 24 minutes
6000 x 4500	77.2 MB	5	15	516	10 hours, 32 minutes
4000 x 3000	34.3 MB	3:20	13:20	457	9 hours, 14 minutes
2000 x 1500	17 MB	1:40	11:40	400	8 hours

Copy setup for 30x40 original, 2 Arri 1000 watt tungsten Fresnel instruments @ 12 feet. Illumination at copy stage 1647 lux (153 foot-candles) Camera settings: aperture f11, ISO 400, line time 1/15 second

Pixel dimensions	File size	Scan time in minute	Total time under lights in minutes	Total exposure in lux/hours	Equivalent gallery time at 50 lux
12000 x 9000	309. MB	13:20	28:20	777	15 hours, 30 minutes
8000 x 6000	137.3 MB	8:53	23:53	646	13hours
6000 x 4500	77.2 MB	6:40	21:40	593	12 hours
4000 x 3000	34.3 MB	4:27	19:27	534	10 hours, 40 minutes
2000 x 1500	17 MB	3:20	18:20	502	10 hours

are not. If total time under the lights were 20 minutes, these exposures would result:

Lux	Foot Candles	Time Under Lights	Total Lux/ Hours	Equivalent at 50 Lux
3000	280	20 minutes	900	18 hours
4000	372	20 minutes	1,333	26hrs, 40 min
5000	465	20 minutes	1,667	33hrs, 30 min

I haven't encountered any copy lighting higher than 5000 lux. Erik Landsberg at MOMA provided meter readings at one of their reprographic workstations, which translates to about 4600 lux. His estimate of 20 minutes as a maximum exposure time agreed with mine. The MOMA photographers are careful to minimize exposure, and cover the artwork during any interruption to the capture process.[12]

Twenty minutes exposure to 5000 lux would result in the equivalent of about four days in a 50 lux gallery. This is probably close to a maximum exposure for copywork. With the shorter scan times at 5000 lux, an experienced photographer probably wouldn't need 20 minutes with most subjects. In practice, I think of most scans in terms of using up a day or two of exhibition time, or four days maximum.

Minimizing Exposure

It's worth remembering that not all, or even most works on paper fall into Category 1. Photographic prints for example include very durable toned black-and-white silver gelatin prints and very fugitive albumen and color prints. Photographers should be aware of the characteristics of the material, and adjust their techniques accordingly.

Some expert operators have developed low-light methods involving multiple scans of the same image. Calculations would reveal whether this is advantageous, given reciprocity. Because some of the more vulnerable artifacts may not have a lot of important shadow area, the ISO setting might be higher than one would choose for a rich full-scale subject, allowing either lower light levels or shorter scans. Some operations don't require full light. During focusing and other adjustments, scrims can be introduced between the lights and the subjects, or the lights dimmed if equipped with rheostats.[13] During interruptions, the piece can be covered or shaded.

During digital capture, lenses can be used at wider apertures than customary when shooting film. Scan backs feature electronic focusing aids which allow much more precise focus than the traditional method. Because high-resolution digital images can be viewed at extreme magnification on screen, imprecise focus or imperfect alignment of subject, lens and camera are easily discovered, as are lens flaws. Once these are addressed,[14] the photographer can work comfortably at wider apertures rather than stopping down to gain depth of field.[15]

Light Sources and UV

Scan backs can be used with any continuous light, including daylight, tungsten, florescent, HMI[16] or even mixed sources as long as the light is homogenous across the subject. In the studio, they work particularly well with special full-spectrum florescents and HMIs. Special tungsten lights from Tarsia Technical Industries (TTI) are being used by MOMA and other institutions with good results. These have dichroic reflectors that pass much of the infrared out the back and away from the subject. This keeps direct heat away from the subject, but raises the ambient temperature, so effective air-conditioning is necessary. Conventional tungsten lights can be used with scan backs, the heat being the primary danger to delicate subjects. Ordinary tungsten sources produce more heat than visible light and their spectrum drops off at the blue to violet end. To achieve a neutral color balance, digital cameras have to boost the output of the blue sensors. They do this quite effectively, but this increased gain tends to increase noise in the blue channel, especially at the marginal light levels that might be used for light-sensitive subjects. However, some photographers prefer the sharper, more focused beam of tungsten lights for some kinds of work. Tungsten lamps in general have an UV component of about 70 miliWatt/lumens, and museums don't generally consider UV filtration necessary for them. Art Preservation Services of New York recommends 70 miliWatt/lumens as a threshold, higher levels requiring filtration.[17]

HMI's produce ample amounts of daylight-balance light and are favored by some commercial digital studios. They appear to be fairly rich sources of both infrared and ultraviolet[18], but at least one museum is using them for digital capture of paintings and sculpture, if not works on paper. According to Brad Flowers of the Dallas Museum of Art, a pair of its HMI's in soft boxes at 7 to 8 feet distance produce UV readings of 160 miliWatt/lumens. Bare at 4 feet, the UV level was 200 miliWatt/lumens. HMI sources probably aren't appropriate for sensitive works on paper, but might be useful for large scale works of more durable materials, which are difficult to light brightly enough for digital capture.

New florescent designs have emerged to meet the needs of video and digital studios, and seem to be becoming the light of choice for most scan back photography. They're efficient, produce little heat and emit a spectrum well-suited to the sensitivity of CCDs. Because florescent tubes use ultraviolet internally to excite the phosphors that produce the visible light, they have been suspected as sources of damaging UV. Many museums require UV filters over any florescent gallery lighting. Most of the florescent studio lights on the market use Osram tubes. David Christensen of North Light Products has provided data for the Osram tubes used in his equipment, and the UV component appears to be no higher than that of tungsten lights. Measurements taken at the San Francisco Museum of Modern Art Conservation Lab seem to bear this out.

Recently, Thomas Palmer scanned a number of Ansel Adams photographs at the San Francisco Museum of Modern Art using a Dicomed scan back with a pair of 200-watt Balcar florescents with bare Osram tubes. Conservator Theresa Andrews took measurements at the copy stage, recording a luminance of 250 foot-candles (2688 lux) with an ultraviolet component of 50 miliWatt/lumens. Temperature was also monitored, concern about heat effects being one of the main reasons for using florescent lighting. Scans ranged from three minute to nine-and-a-half minutes. The Balcar lights feature dimmers, variable down to 25 percent of full output, and this feature were used to minimize exposure during focusing and

other operations not requiring full light. Dimmed, the lights produced 15 miliWatt/lumens in the ultraviolet region.

Foot-candles: 250
lux: 2688
UV: .....50 mW/lumens
Scan time (max): .....9.5 minutes
lux/hours: 425.6
Equivalent exposure @ 50 lux (scan only): 7 hours, 55 minutes

Last year, during another publishing project at SFMOMA, Robert Henessey digitized Carlton Watkins albumen prints using a Better Light scan back with two 120-watt florescent K-Lites. With 400 foot-candles (4300 lux) at the copy stage, UV levels were 100 milWatt/lumens. Scan times were only three to three-and-a-half minutes.

Foot-candles: 400
lux: 4300
UV: .....100 mW/lumens
Scan time (max): .....3.5 minutes
lux/hours: 251
Equivalent exposure @ 50 lux (scan only): 4 hours, 40 minutes

The Balcar UV readings were less than the tungsten average, the K-Lites a little higher, but not by a large amount. The florescent lights being used for digital photography don't seem to be strong sources of ultraviolet, the levels being in the vicinity of tungsten readings. With filtration, the amounts should be reducible to nearly zero. The lights made by North Light Products incorporate Plexiglas panels in front of the tubes, which the manufacturer will fit with UF3 ultraviolet reducing Plexiglas on request. Other makes could be fitted with UF3 shields, with care taken to allow space for ventilation between the tubes and the shield if the lights aren't fan-cooled like the North Light designs. The TTI tungsten lights at MOMA also have UV-absorbing Plexiglas shields to eliminate the moderate amount of UV, and further reduce heat.[19]

Karen Colby assumes the absence of UV in her calculations. Timothy Vitale, on the other hand, didn't consider the ultraviolet component of cold cathode scanner lamps (assumed to be between 0.7 percent and 2.4 percent of total output like other florescent sources) to be a significant factor.[20] Nor did Stefan Michalski in reference to flash exposure. Both affirm that sensitive materials are more affected by cumulative exposure to visible light than to the small amounts of UV under controlled light levels. The light exposures from scanners and amateur electronic flash are considerably lower, however, than required by scanning cameras. If UV at these levels is a concern, it should be possible to eliminate it with filtration.

Museum Policy for Photography

Part of the rationale for digitizing fragile or sensitive objects is to provide access for study and scholarship without handling or exhibiting the original so often. It might be regrettable if many artifacts become available only in reproduction, but it would unquestionably extend the life of the original, and even if fewer people see the original, the reproduction will be much more widely accessible.

Digital capture should reduce the number of times a piece would have to be re-photographed, as is common now to replace lost, damaged or deteriorated transparencies. Although there's a lot of concern about the permanence of CD-ROM media, the data on them will probably have to be transferred to DVD or whatever becomes the next standard, long before the CDs themselves deteriorate.[21] Digital image archives will have to be maintained and migrated to each new storage medium or file format, but if this is done, the images should suffer no loss and be useful for a long time.

Scanning cameras can record a greater range of values than can be displayed by any current means, recording brightness ranges of 10 or more stops. Color transparency film has a range of 5 or 6 stops and nothing else, print or video display, comes close to that. The top-end models exceed the resolving power of many lenses. Put another way, image quality is better than we can fully use at present and if deep-bit files of sufficient resolution are archived, they should provide adequate documentation for many years.[22]

Digital capture offers some attractive advantages, and a number of museums have already commenced digitization projects, regarding the light exposure involved to be justified by the benefits. If other studies support the conclusion that it involves the equivalent of a few days of exhibition time at low-light levels, it should be easier to make informed decisions about digitizing collections and individual pieces, and to monitor total exposure more effectively.

Special thanks to the following individuals, companies and institutions:

Jill Sterrett and *Theresa Andrews*
Conservators
San Francisco Museum of Modern Art

Erik Landsberg
Manager of Imaging Technology Development
Museum of Modern Art, New York

Tom Jenkins and *Brad Flowers*
Photographers
Dallas Museum of Art

Michael Collette, *Larry Guyer*, and *Robin Meyers*
Better Light, Inc.*
1200 Industrial Road, Studio 17
San Carlos, CA 94070-4129
Phone 650/631-3680 • Fax 650/631-2915
www.betterlight.com

David Christensen
North Light Products, Inc.*
2487 Spring Street, #2
Redwood City, CA 94063
Phone 415/366-5483 • Fax 415/366-1676
www.northlightproducts.com

*Better Light, North Light and Balcar are distributed by
Calumet Digital Solutions
http://www.calumetdigital.com/

Other sources and resources:

Garry Thompson. *The Museum Environment*. Butterworth-Heinemann, Butterworth-Heinemann Series in Conservation and Museology, 1994

Henry Wilhelm and Carol Brower. *The Permanence and Care of Color Photographs*. Preservation Publishing Company, 1993

Donald P. D'Amato and Rex C. Klopfenstein, March 1996, *Requirements and Options for the Digitization of the Illustration Collections of the National Museum of Natural History* at *http://www.nmnh.si.edu/cris/techrpts/imagopts/index.html*

Conservation Online Web site at *http://palimpsest.stanford.edu/*

Academic Imaging website at *http://www.academicimaging.com./*

Image Permanence Institute Web site at *http://www.rit.edu/~661www1/sub_pages/frameset2.html*

Ben Blackwell
Photographer
UC Berkeley Art Museum & Pacific Film Archive

Ben Blackwell is the staff photographer at the Berkeley Art Museum/Pacific Film Archive, and maintains a freelance photography business specializing in fine art reproduction for clients that include major museums, publishers, galleries and collectors.

[1]Stefan Michalski, Light Effects from Flashbulbs and Copiers, Abbey Journal, Volume 20, Number 6, Nov 1996, posted at http://palimpsest.stanford.edu/byorg/abbey/an/an20/an20-6/an20-607.html
[2]Saunders and Kirby, "Light induced damage: investigating the reciprocity principal," ICOM-CC Preprints, p. 87-90
[3]One foot-candle is defined as the intensity of light that falls upon a one foot square surface illuminated by a light source that equals one candle power, or candela. Lux is the metric measurement: the light on a one square meter surface, one meter away from a source of 1 candela. One lux equals 0.0929 foot candles, or roughly 1/10 of a foot-candle. Therefore, 5 foot-candles is about 50 lux.
[4]A bare 60-watt bulb in a large room produces 50 lux on a surface about 3 1/2 feet away; or aimed in a reflector desk lamp, about 6 feet distant. Colby notes that older viewers may have trouble seeing artwork exhibited at 50 lux.
[5]Karen M. Colby, "A Suggested Exhibition/Exposure Policy for Works of Art on Paper" available at http://www.lightresource.com/policy1.html
[6]The British Blue Wool Standards consist of eight dyed wool test specimens ranging from very fugitive (ISO 1) to very resistant (ISO 8). Each numbered specimen requires twice the exposure to exhibit the same degree of fading as the previous number. Each of Colby's categories incorporates three of the ISO levels and is based on the middle, i.e. Colby's Category 1 covers ISO categories 1, 2 and 3, and her recommendations for that category are based on objects with a light sensitivity of ISO 2.
[7]In practice, most artifiacts aren't necessarily exhibited every year. 12 weeks exhibition every three years would result in the same exposure as 4 weeks every year.
[8]Light exposure in scanners has been well-addressed by Timothy Vitale's study: Light Levels Used in Modern Flatbed Scanners, found at http://www.rig.org/preserv/diginews/diginews2-5.html#technical
[9]There are other designs of digital cameras. Some, like the Jenoptik eyelike cameras, are being used for art reproduction.
[10]Scan backs from other manufacturers, Dicomed, Phase One or Jobo, may vary in their scan times, but because the total time under the lights varies so much due to other factors, relatively small differences in scan time needn't be considered.
[11]Noise in digital images consists of anomalous pixels, usually appearing as random flecks of color in dark areas and looking something like film grain.
[12]MOMA was one of the pioneers in adopting this technology wholeheartedly, and their ambitious and well-planned digitization project has been a model for other institutions. See *Digitizing Photographic Collections: A Case Study at the Museum of Modern Art*, NY, Colet, Keller, and Landsberg, Spectra,Winter 1997.
[13]It's not advisable to turn florescents off and on, because they require time to warm up and stabilize.
[14]The Zig-Align, a simple and ingenious mirror device is a useful aid in aligning camera, lens and copy stage. Contact William Zigler, P.O. Box 765, Menlo Park, CA 94026, 650/342-3704
[15]Close examination of these digital images forcefully illustrates the fact that depth of field is an illusion, based on the size of blob the eye will accept as a point (the circle of confusion.) As the lens aperture gets smaller, out-of-focus blobs get smaller, but there's still only one plane of precise focus.
[16]Hydragyrum Medium Arc-length Iodide lights produce daylight-balanced illumination in an instrument resembling a tungsten fresnel attached by cable to a large ballast unit.
[17]See the Elsec UV Monitor Type 2 instruction manual, Art Preservation Services, 539 East 81st Street, New York
[18]Arc sources, like HMI and Xenon, have UV in their spectrum, some of which is blocked by the lens in front of the bulb.
[19]Colet, Keller and Landsberg, *Digitizing Photographic Collections: A Case Study at the Museum of Modern Art*, NY, Spectra, Winter 1997
[20]Timothy Vitale, *Light Levels Used in Modern Flatbed Scanners*, page 12
[21]Concern over CD and other media longevity is pertinent in that the most dependable should be identified and adopted.
[22]Most scan backs operate at 14 bits per channel, recording 16384 levels of each primary color. Present print and display equipment uses only 8 bits per channel, for 256 levels. Bruce Fraser, among others, advocates archiving deep-bit files against future improvements in output technology.

MCN SPECTRA

A Quarterly Publication of the Museum Computer Network

❑ Renewal ❑ New Subscription

Date: _____

Institution: _____

Office/Division:_____

Mailing Address:_____

City: _____

State/Province: _____

Zip Code: _____

Country: _____

Contact Name: _____

Telephone: ( ) _____

Fax: ( ) _____

E-mail: _____

Subscription to the *MCN Spectra* is available to libraries only.

❑	*Spectra* Subscription	$ 75.00
❑	Postage - Outside U.S. and Canada	$ 10.00
	Total:	$ _____

PAYMENT INFORMATION

❑ Personal Check Check Number _____

❑ Company/Institutional Check Check Number _____

❑ Credit Card ❑ Visa ❑ MasterCard ❑ American Express

Card Number _____ Expiration Date _____

Signature _____

FOR MCN OFFICE USE ONLY

Date Received: _____ Date Input: _____ Date Acknowledged: _____

Museum Computer Network
1550 S. Coast Highway, Suite 201 • Laguna Beach, California 92651
(877) 626-3800 • info@mcn.edu • www.mcn.edu

Conservation Issues in Digital Imaging

Norbert Lossau, Martin Liebetruth
Göttingen State and University Library / Digitization Center, Germany

Preamble

In this article, we will offer you a deep insight into the conception, purchase and installation of a high-quality digitization workstation at the Göttingen State and University Library. The article provides detailed information on our policies of handling irreplaceable objects from our collection featuring our recent project: "Digitization of the Göttingen Gutenberg Bible." This bible is one of the few complete vellum copies worldwide, and as a result of our project, it's now available for browsing via the Internet (*www.gutenbergdigital.de*). The project-oriented approach to the topic of "Conservation and Digitization" as exemplified by this article demands complementary research on a more abstract level. However, the strong pragmatic focus may be of interest to those of our readers who are currently implementing or contemplating the implementation of digital access projects for valuable and fragile holdings of their institutions.

Introduction

"Conservation" and "Digitization" represent key terms from two different worlds. Conservation has overtones of long-standing tradition, carefully applied to holdings of honorable institutions like museums, libraries or galleries to ensure the longevity of their treasures. Digitization has the flavor of new technology, and although it is becoming a more widely accepted strategy for improving access to a lot of different resources, it still comes with imponderabilities in terms of digital asset lifespan, authenticity, etc.

Every digital imaging project deals with issues of preservation in a twofold sense—the preservation of the original and the preservation of the surrogate. Preservation of the original in the context of digital technology does not really raise entirely new issues. At first sight, digital conversion is simply an additional way of reproducing original material and in a lot of respects raises the same issues as microfilming or other analog photography. Going into the details, however, we detect a number of issues specific to digital photography, e.g. particular lighting needs and long exposure times.

Most of the preparatory steps for the conversion procedure are not new. Assessing the condition of the original with respect to possible strains during the conversion process in terms of fragility, light sensitivity, binding structures etc. are not unique to a digitization workflow. Any conversion process has the overall goal to protect the original and minimize any physical handling. Furthermore, any conversion has to be performed with the intent of creating a (near) faithful reproduction of the original in order to avoid the necessity of repeating the procedure.

Although objections from curators tend to focus on the preparation and conversion process, more vital concerns in respect to digitization arise from the uncertainties related to long-term maintenance of digital data, issues of digital image authenticity and uncontrollable distribution via electronic media (copyright issues).

Long-term maintenance of digital data belongs to the broad field of surrogate preservation and there is a long and open-ended discussion about it. We acknowledge the issues raised in the discussion, but don't want to go into details in our article. From our point of view, the application of standards is the only way to minimize potential risks for the longevity of digital data. In the given context, standards mean: standard graphic file formats (uncompressed TIFF for tonal images), standard storage media (CD-R, ISO 9660), platform independent color profiles (ICC profiles) and standard (open) metadata description (ASCII, SGML, XML).

Authenticity of the original and its distribution can be secured on different levels and will not be a topic in our article. Beyond watermarking and other, more sophisticated "tracking mechanisms" of online image access, we rely simply on the low quality of our online display images, which will not allow for inappropriate, illegal reproduction.

Conservation Issues and Digitization

Conservation in general has its own more or less rich tradition in our public "collecting communities" (museums, libraries, archives, galleries), and every community has its more or less elaborated common guidelines, often supplemented—or superceded—by a set of rules at individual institutions. For libraries, the recent publication *Principles for the Care and Handling of Library Material* of the International Federation of Library Associations and Institutions (IFLA) provides a comprehensive overview about the impact of environmental factors on collection materials.

Some of these environmental factors play an important role in digital imaging projects. Apart from the issues involved in physically handling fragile objects, other widely known problems are heat, climatic conditions and light (types of lighting, light and UV level). While most of these issues have already been adressed for microfilming by our colleagues from photography units, new concerns arise with the advent of digital technology. What type of digital camera is best suited for capturing the objects I am responsible for? What level of quality can I expect from the digital conversion process? What kind of compromise between object protection and image quality do I have to make? Which type of lighting can I use with which camera? What about exposure time, UV levels, light levels and temper-

ature increases? Is there sufficient depth of focus in digital photography, or to what degree do you need to flatten the original (page)? Are there special constructions to support contact-free scanning (cradle, *no* glass plate)? This whole range of questions could be continued, but for the time being, let us just add one more issue: reasonable cost for the right digital equipment to answer all your needs.

There is not an abundance of guides or literature in this special field examining these issues from the library point of view. Answers to these questions can be found only in special discussion groups or via e-mail lists. IMAGELIB, hosted by the University of Arizona and maintained by Stuart Glogoff, is such a forum for researchers, practitioners and project managers, and has proven to be a valuable information resource for our work in Göttingen. Recently, various organizations have increased their efforts to address these questions. The successful September 1998 conference of the Research Libraries Group (RLG) and the National Preservation Office (NPO, UK) in Warwick was a promising start in this area.

A lot of expertise in this field is widespread over different countries, but there are only documentations for individual institutional efforts, e.g. as provided by the Library of Congress or the Bodleian Library at the University of Oxford. Some helpful advice may be gleaned from two publications for digitization projects edited at Cornell University Library, and in the joint study of the Visual Arts Data Service (VADS) and TASi (Technical Advisory Service for Images) in the United Kingdom.

The following information describes the route taken at the Digitization Center of the Göttingen State and University Library (GDZ) in designing, purchasing and implementing a special color digitization unit for the collections of the library. The digitization of the Göttingen Gutenberg Bible will be examined in detail, illustrating the daily workflow in respect to conservation as well as image quality assurance and storage procedures.

Case Study: The Digitization of the Göttingen Gutenberg Bible

Some Reflections Concerning Cooperation of Digitization Units and Curators / Preservation Librarians

Planning and implementing a digital conversion project will challenge you to make numerous decisions in the field of technology. However, try not to get lost in the technological issues. You are not on an island where you have to face the adverseness of the daily project on your own. The best way of getting to know the treasures of your institution is by way of the curators who devote their time to presenting and preserving the objects in the best way they can. They know the history of the object you have to digitize the best. They are able to assess the actual condition of the object and judge its fragility. They can make recommendations how it will best withstand the physical stress of the conversion process. Don't try to convince them that there is no physical stress for the original—there always is—but you can minimize it with the help of curators, and by adapting your equipment to the requirements of the objects at hand.

They will not agree to disbind or put at risk an original which has survived several hundred years in its genuine leather binding just because of your difficulties in digitizing a curved vellum page. There are cradle solutions with various flexibility—some already in use for microfilming projects, some with new features specifically developed for digital photography. Line array digital cameras often have a very long exposure time, and you have to minimize the strain on the object by choosing a camera with short exposure times. This enumeration is only a very limited selection of questions and problems you have to deal with when embarking on your digital conversion activities, but it illustrates that you only have a real chance to successfully implement your projects when you reach agreements with the curators of the original object. Moreover, finding a common level of understanding with your preservation librarians and curators gives you the chance to learn more about the original itself. Curators also have insights into the access expectations of potential users. Measuring these expectations to the technological feasibility can result in a digital surrogate which meets a wide range of interests. The information curators provide can help you find the best way of digitizing and presenting an object according to its value.

The Göttingen Gutenberg Bible, Vol. 1, fol. 5r: *(Beginning of the Book of Genesis)*

General Framework

The Digitization of the Göttingen Gutenberg Bible is an example of successful cooperation within the library. Dr. Helmut Rohlfing, head of the Department of Manuscripts and Early Imprints and the Digitization Center, upon recognizing the strength of each partner, found a very constructive way to set up and implement the digitization of this extraordinary object, one of the rare complete vellum copies of the Gutenberg Bible.

Technical Infrastructure at the Göttingen Digitization Center

To understand the technological background of the project, the following paragraph provides a brief introduction to the Göttingen Digitization Center (GDZ). In 1997, the Göttingen State and University Library started to set up one of two national Supply Centers for Retrospective Digitization of Library Material, funded by the German Research Foundation (Deutsche Forschungsgemeinschaft, DFG).

The Lower Saxony State Ministry of Education and Culture provided the GDZ with a substantial grant to build a high-quality technological infrastructure.

Beyond two bound volume scanners (Zeutschel OS 7000 and Minolta PS 7000), a special grant of the State Ministry allowed the set up of a special color digitization unit at the GDZ in 1999. Following the international discussion in the field of digital imaging during the last two to three years, we decided in favor of direct capture for the digitization of our library treasures. The special digitization lab is located in a separate room at the GDZ premises within the historic building of Göttingen Library. The historic building provides good climatic conditions for exposing old sensitive book objects—not too hot and not too dry—and it also has a (near) vibration-free floor.

The Original - Göttingen's Copy of the Gutenberg Bible

The 1,282 pages of the Göttingen copy of the Gutenberg Bible are divided into two volumes. The text was printed on vellum, whereas column headings are handwritten in red missal script. The capital letters were decorated lavishly with burnished gold, colors and foliage. The beginning of the Bible and the first pages of each individual book of the Bible are decorated with foliage depicting akanthus, ferns or briers. All in all, there are 88 pages with illuminations.

The size of the pages is 29.4 cm by 40.1 cm. A considerable number of the pages are not completely plane because vellum changes its shape when exposed to changes of relative humidity. The physical condition of the vellum pages is still excellent, without any instances of brittleness or fragility. The binding of the two volumes derives from the 16th century. Before the beginning of the conversion process, the binding was treated with special wax to allow for a greater flexibility of the leather. Both volumes could easily be opened at an angle of 1400.

Digitization Equipment and Computer Hard-/Software Environment

Basic components of our color digitization workstation are:

- Digital camera back
- Analog camera (4x5)
- Light sources
- Working place with moveable cradle
- Measuring instrument for temperature, light level (lux) and UV radiation

Hardware components include a computer workstation including a 21-inch monitor to evaluate the quality of the images, mass storage devices to handle the huge amount of data, and a CD-ROM burner to create backup and archival copies.

Special software is used to drive the digital camera, as well as for device calibration, color management and image processing.

Digital Camera

At the time of our selection process, two main types of digital cameras, differentiated by detector arrangement, were available on the market:

1. Area array systems
2. Line array systems

Area array cameras allow fast shooting (1-shot camera), resulting in short exposure times, and they support flash photography. The resolution varies, e.g. from 2048 by 2048 pixels (Eyelike) up to values of 4096 X 4096 (Dicomed Big Shot 4000). The resulting file sizes range from 12 MB to 50 MB.

Line array cameras need more time to move the CCD across the back of a camera to collect the incoming light, but they produce high-resolution images. The basic technology may be compared to a miniature flatbed scanner. To ensure high-image quality, constant, flicker-free lighting is needed. The object cannot be removed during a longer exposure time (e.g. to prepare the next scan). The resulting image size is considerably larger than with area array cameras and can measure up to several hundred MB. Files created with line array camera backs contain more information and feature more detail.

The first research into extending the technical infrastructure of the GDZ in the direction of color digital photography was conducted in 1998. Recognizing the difficulties in gaining substantial information about products (see Conservation Issues and Digitization on page 24), we went the classical route of gathering information in the field of digital technology: research on the Internet, learning from various other digitization projects— where documentation was available—mainly in the U.S. (e.g. at the University of Virginia, Library of Congress, UC Berkeley, Harvard) and studying mailing lists (e.g. IMAGELIB). The already-mentioned Warwick Conference in September 1998 was extremely helpful, especially the comprehensive overview from Stephen Chapman (Harvard University) about "Guidelines for Image Capture."

Gaining general product information about digital cameras, available in vendor literature, is one thing. To enrich this general information with special knowledge about their use in libraries, museums or archives is much more complicated. We decided to consider only the digital camera backs we could test in-house on objects from the library's collection. Recognizing the special requirements of our older library material led us to insist on in-house test runs.

We searched for a digital camera not only for one specific project, although the digitization of our Gutenberg Bible had the highest priority. After this project, the camera would be used for other objects (incunabula, maps etc.). In other words, we were looking for a camera with the ability to handle materials of different sizes by providing high resolution for each of the potential objects. Three-dimensional photography was not an issue for us at the time.

Starting from these general considerations, we tried to define a list of requirements. Our main priorities:

- **High resolution:** We benchmarked the scanning resolution for our digital captures at a minimum of 300 ppi. Taking into consideration our goal of capturing objects of varying sizes (e.g. a page of the Gutenberg
Bible has a size of 29.4 cm by 40.1. cm), and to keep the option for different output versions open (like the display on the Web or high-quality facsimile reprint of the digital surrogate), at least a 6000 X 6000 pixel array seemed neccessary.
- **Exposure time:** As short as possible, while keeping the resolution as high as possible.
- **Lighting:** Flash photography is not an option from the conservation

point of view, so we started looking for a source of constant light without noticeable increase in temperature and with a full-color daylight spectrum (color temperature of about 5500 Kelvin). UV radiation is a concern, but may be reduced by using UV filters.

- **Scanning software:** Integrated camera/software solution, support of ICC profiles.

Based on these requirements, we created the following list of camera products, annotated with basic technical information:

- **Better Light** (Better Light) digital scanning back, 8000 x 10640 pixels, maximum 488 MB RGB image size, Kodak trilinear color CCD technology, line array.
- **Dicomed Big Shot 4000** (Dicomed), digital scan back, 4096 x 4096 pixels, up to 48 MB image size, area array.
- **Eyelike** (Jenoptik) digital camera system, 2048 x 2048 pixels, maximum 108 MB image size, CCD area array technology.
- **Kontron Progres 3012** (Jenoptik), digital camera system, various resolutions up to 4608 x 3480 pixels, image size 48 MB, for high resolution a moveable area array chip (Piezo technology).
- **MARC II** (Lenz brothers, Munich), digital camera system designed to capture images up to 15,000 X 15,000 pixels. One system we know about is being used by an imaging service bureau in the United States. Due to the potentially tremendous image file size, the company used only a pixel array of 7450 X 5600 for its production. The system uses a combination of array and linear array cameras.
- **Phase One Powerphase**, (Phase One) digital scanning back, 6000 X 8400 pixels, maximum 144 MB image size, line array.
- **PictureGate 8000**, (Anagramm) digital scanning back; 8000 X 9700 pixels, maximum 444 MB RGB image size, trilinear color CCD technology, line array.

The Photokina '98 in Cologne offered the chance for us to get in touch with some of the vendors for the digital cameras listed above. We succeeded in conducting tests with three of these cameras: Kontron Progess 3012, Phase One Powerphase and PictureGate 8000. The MARC II had been produced (at this time) only in a very limited number, and we were not able to obtain a camera for a test in our library. In addition, the camera software available in Germany was still a prototype and it would have been our task to finance a suitable software solution. We had difficulties finding competent service vendors for the other products in Germany, so we limited our evaluation to three scanning backs.

Tests carried out at the GDZ indicated a long exposure time of three to five minutes for the Kontron and Phase One Powerphase camera systems. Despite the area array technology of the Kontron, the use of high resolutions with this system requires a lighting setup comparable to a line array system. Both from the conservation point of view and under performance (production rate) aspects, three to five minutes per scan itself was not acceptable. In addition, the Kontron revealed some slight shifts in color fidelity which might have been induced not by the camera itself, but by the lighting environment. It should be stressed that our tests were performed in 1998 and due to further developments of the products, a current attempt may modify previous results.

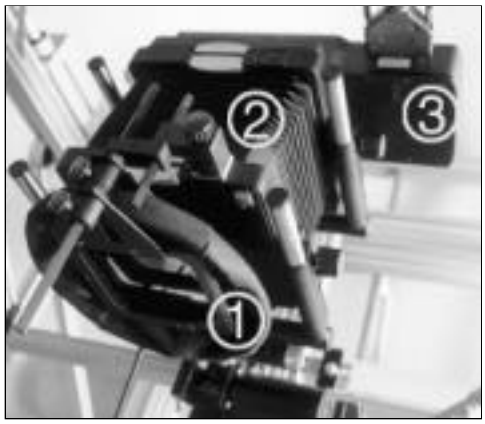
The Anagramm PictureGate 8000 digital scan back provides a maximum resolution of 8000 X 9700 pixels, which at that time was only exceeded by the MARC II and Better Light systems. The file size amounts to 444 MB per scan in 48-bit mode (222 MB for 24-bit mode). Scanning time depends on the image quality, the lighting and the lens aperture. The PictureGate 8000 outperformed the other cameras in almost every regard. It was able to capture images at the highest resolution of 8000 X 9700 pixels and keep exposure times under 120 seconds for file sizes of more than 300 MB. Good tone, color and image quality convinced us to vote for this scanning back. Camera back and software gave the impression of a well-rounded system. It should be mentioned that we were allowed to test the system for almost one week. The relatively high price, about $40,000 U.S., seemed to be justified, and included scanning software and color management as well as one year of software upgrades.

Due to the large file size of the images, we needed a powerful computer with a large internal hard disk (9 GB), an additional external hard-disk system (140 GB) and 756 MB RAM for image processing.

Analog Camera

As an analog camera for our PictureGate 8000 scan back, we use a Linhof Kardan E large format (4x5) monorail camera, equipped with a Rodenstock Apo-Sironar N (5.6/135mm) lens.

The use of an analog camera offers some of the capabilities of traditional photography. For example, we were able to be in focus over the surface of the whole page by closing the lens aperture in spite of the fact that the vellum was not entirely plane. This was the tricky part of the shoot, because the more you close the lens aperture, the more light the camera needs. It took some time for us to find the optimum settings.



1) Rodenstock Apo-Sironar-N high resolution lens (5.6 / 135 mm), 2) Linhof Kardan-E large format (4x5) camera, 3) PictureGate 8000 scan back

Light Sources

Our decision to use a line array scan back led to some distinct demands on the light sources. Longer exposure time requires constant, flicker-free lighting. We conducted a test of different manufacturer's products in-house, which revealed that although the cage of the lamps differed, the lightbulbs remained the same. Our vendors recommended Osram Dulux lamps with a daylight spectrum (color temperature of 5400 Kelvin) and low UV radiation.

We decided to purchase Efpe Design lamps with 6 Osram Dulux tubes of 55 watts each, producing 330 watts altogether. Two cowls (each with 6 Osram Dulux tubes) were needed to illuminate the scan area in a well-balanced manner. The Efpe Design lamps can be dimmed to lower the light levels during set up.

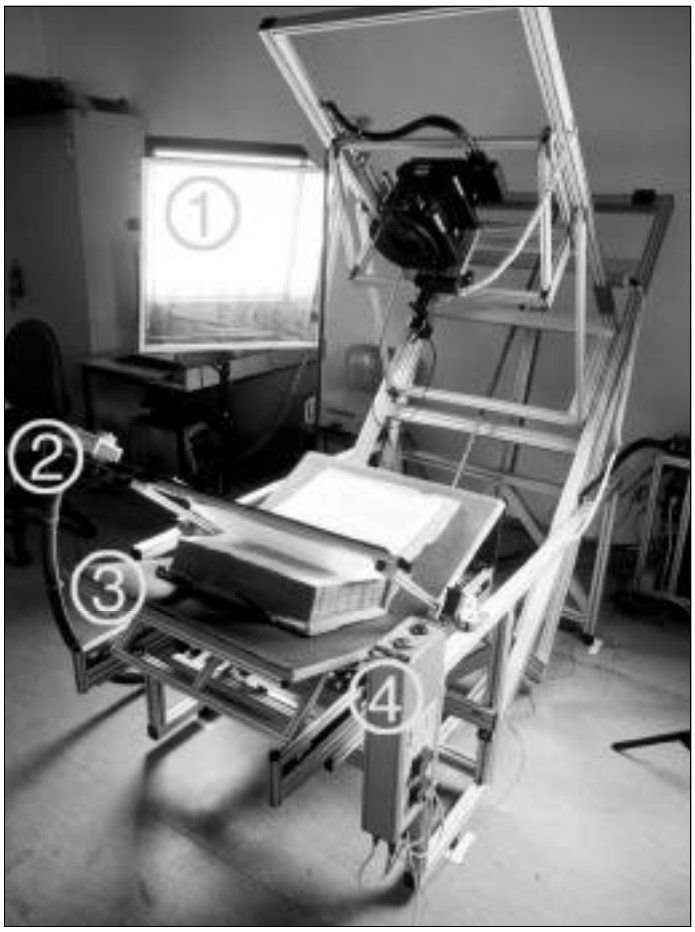
To minimize UV radiation, additional foil was used as a UV filter. We managed to reduce the UV radiation to 10 µW/Lumen as compared to the 90 µW/Lumen of a neon lamp or the 1300 µW/Lumen of a cloudy sky. Inside shutters were used to eliminate direct sunlight.

Each page of the bible was exposed to the 9000 lux of light for a maximum of 10 minutes, which corresponds to 1500 lux-hours (9000 / 6 = 1500 lux-hours). In an exhibition setting, the exposure is 50 lux (seven-hour day, six-day week). This adds up to 2100 lux-hours in one week (50*7*6 = 2100). Keeping in mind the cumulative nature of light (the so-called "Reciprocity Law"), the exposure of a single page during the digitization corresponds to about four days in an exhibition.

From the conservation point of view, this value seemed to be acceptable. However, during the whole scanning process, we considered to optimize the workflow to guarantee a minimum of light exposure. Today there may be other lighting solutions which generate lower heat, e.g. so-called "plasma lights."

Working Place (Digitization Desk)

Another key issue from the conservation point of view is to protect the object against any physical damage. Finding a suitable solution to this problem took a long time during our planning process. Based on our experience with cradles at the GDZ, we decided at a rather early point to look for a custom-designed solution. A sine qua non was the support of contact-free scanning (that means *no* glass plate to hold pages plane). In addition to this requirement, we wanted a solution which would enable restricted opening (an opening angle of 1800 is not realistic for the majority of our older library material). Furthermore, it should feature a split level of the cradle to take care of the book spine and a cradle plate that would allow the scanning of objects of different size up to a minimum of A2.



1) Efpe design area lamp, 2) Perforated vacuum bar to fix the pages, 3) Adjustable book cradle, 4) Control panel

In January 1999, we had the opportunity to visit Manfred Mayer in Graz, Austria. We had heard about a custom-designed cradle solution at the University Library of Graz at several conferences and wanted to discuss the details on location. Mayer is the head conservator at the University Library plus an excellent engineer, and thus uniquely qualified to create a technical solution for this task. In 1997, Mayer had already designed a versatile digitization desk, which meets all requirements as described above.

The desk/stand is made of aluminum, which ensures the longevity of the construction and remains flexible enough for changes. The camera is positioned at the top of the stand. A moveable cradle with split levels is the center of the desk and allows the scanning of large-size objects like maps or plates up to a total size of 95 x 65 cm (having both levels horizontally plane). Books can be opened at different angles and manually adjusted.

The solution to hold pages or leafs flat is ingenious. Mayer applied a perforated vacuum bar that secures the page of a book or a leaf by fine-tunable low pressure while the book lies on the cradle. The construction of the cradle, vacuum bar and camera mounting is designed to keep the axis of the camera always vertical to the page you wish to scan. Another amazing feature is the installation of a laser beam at the axis of the book cradle to facilitate the positioning of the book. Image size and resolution can be altered easily by changing the distance between camera and book page. All operations of adjusting the camera can be performed automatically from a panel on the desk or alternatively from a remote panel, which may be located near the computer desktop. Markers on the desk allow it to resume positions that have been found ideal for different sizes or types of objects.

Computer Hard-/Software Environment

Macintosh computers are well-known in the community of graphic design for the processing of digital color data. We chose Apple MacOS as the suitable platform to handle the tremendous mass of data we expected from our Gutenberg digitization project (during the project we captured 380 MB archival files for each of the 88 illuminated pages, and 180 MB files for the text pages). In addition, the driver software for our scan back was only available for the MacOS platform at this time. We selected a PowerMac G3 (OS 8.6) with a 333 Mhz processor and 756 MB RAM. Thanks to Apple Germany, we were able to expand our computer platform with a second PowerMac G3, 450 Mhz processor and 1 GB RAM during the course of the project. Included was an Apple 21-inch self-calibrating display. To store the image files during a day's scanning, we used an external hard-disk array with a capacity of 140 GB.

The PictureGate 8000 scan back is delivered with its own driver software. It allows you to determine settings, such as exposure time, quality and curves adjustment. The software supports the embedding

of ICC profiles into the image files and has saving options for TIFF and the native PhotoShop format.

Digitization Process Calibration

Color calibration is the nuts and bolts of digital photography. Each device of the capturing system has to be calibrated to obtain a (near-true) color reproduction of the original.

Every morning before the beginning of the actual image capture process, we had to calibrate the camera by setting a neutral gray point while using constant lighting environment. This neutral tone is chosen from a gray card supplied by the camera manufacturer. The computer monitor had to be included in the calibration routine. We used the Apple system extension ColorSync, which supports the easy implementation of a color- management system.

Image Capture

After the routine calibration of the image capture components, the actual shooting process could start. We agreed with our colleagues from the department of manuscripts and early imprints to check out the Bible for a daily five-to-six hour scanning process. Carefully, wearing cotton gloves to avoid direct contact with the vellum pages, the Bible was positioned faceup on the book cradle. The volume was opened to a maximum angle of about 120 to 140 degrees to protect the binding. It turned out to be more advantageous to capture only single pages, and we decided to scan first all verso (left-hand) pages of a volume first, then all recto (right-hand) pages.

Once we thought we had found the right positioning, we double-checked it with the laser beam. The laser beam marks the ideal position of the object to ensure optimum sharpness of the image capture. To adjust the opening angle of the volume, a crank handle can be used (maximum alteration 12 cm). The horizontal shift is adjusted by hand (maximum alteration 15 cm). Following the prep work, a color target was placed near the Bible, and the prescan, which takes a maximum of 30 seconds, could be started.

The remote panel allows the user to change the position of the objective to gain the focus. Within a zoom window, a live picture is transmitted directly from the camera to the screen so that the optimized distance can be found interactively. To optimize the color, we determined the white point of the image with the scanning software. The actual image capture process takes about 90 to 120 seconds, depending on the scan area. The captured image was then evaluated on the screen. If the image quality seemed satisfactory, the file was saved as an uncompressed TIFF in 48 bit color depth on the external hard-disk array. The time needed to save the file (90 to 120 seconds for a 200 MB file) could be used to prepare the next page of the Bible. This procedure helped us to minimize the light exposure of each page. On average, we had a production rate of 30 to 40 pages per day (five to six hours).

Conservation and Image Quality Assurance

Every reformatting process requires some monitoring routines to ensure image quality and suitable conservation conditions. As we discussed above, one of the most important issues from the curator's point of view is the climatic environment the book is exposed to during the digitizing process. Normally, the Bible lies in a climatically controlled vault.

To minimize the physical stress for the book, we determined a maximum scanning time of five to six hours per day. A temperature, lux and UV meter was placed near the workstation. Readings taken during the day indicated only an increase of two to maximally three degrees celsius in temperature thanks to the special cold light lamps we used (maybe the German winter helped us to keep the climate reasonable, too). The height of the working space in the historical library building might have had a positive influence on the temperature and other climatic conditions.

During the course of our image capture process, we tried to keep image quality consistent. One of the essential activities was the calibration process described above. We devised a routine to monitor the camera back and the display at regular intervals. Sometimes we made test prints on our calibrated Tectronics laser printer.

Storage

Investing money and time in a digital conversion process only makes sense if you ensure the survival of your data. We will not discuss the life expectancy of the digital surrogates (see Introduction). But, we want to describe our pragmatic approach to ensure the longevity of our digital files. As described above, during the image capture process the archival-quality images were first saved onto the external hard-disk array connected to the PowerMac. With a capacity of 140 GB, this array allowed us to store the output of a few days worth of image capture. Due to the large file sizes, we had to store about five GB to seven GB daily.

These digital master images had to be backed up to prevent data loss in case of any technical problem with the disk array. Our first choice to save the images was a CD-burning device (Plextor plexwriter 8/20) connected directly to the MacOS computer. Although the burning time was rather short, the whole process still remained very time-consuming because the PowerMac is not able to process any other tasks during the burning process. Since you can store three high-resolution images (each 180 MB) on one 650 MB CD-R, and the burning procedure takes approximately eight minutes, you spend about 90 to 120 minutes to save the image production of only one day.

Admittedly, we had underestimated the time it takes to backup our data. If backup cannot be done automatically overnight, you spend about 30 percent of your work time burning CD-Rs. During the final months of the Gutenberg digital project, we could extend out hardware components with a Grundig CD-R jukebox to produce archival copies of the digital masters. The Grundig GMS 1035 has a changeable magazine with 35 CD slots and includes two 4x burners. The jukebox can be operated in batch mode so that the duplication process of 17 CDs is running overnight. The CD-R-jukebox was a valuable addition to our technical infrastructure and resulted in substantial timesavings during production. At the end of the image capture process, we had produced two sets of CD-Rs (about 700 CDs each), containing the complete digital facsimile of our Gutenberg Bible. Each of these sets is stored in a different place.

The "Product(s)" of the Digital Conversion Process

The GDZ follows the strategy of capturing one rich digital master file, then creating derivatives for all other purposes. While applying

this capture principle to all regular original print material simply makes sense, it proves absolutely mandatory for the capture of invaluable, fragile objects. There won't be a second chance to capture them again. As stated above, we rely on standards and decided to save the files as uncompressed TIFFs in 48-bit color depth. Internally, the images are captured in 36-bit mode but can only be saved as a 48-bit file. Alternatively, images also can be saved in 24-bit mode, but our decision was to retain the highest possible color spectrum the camera provides.

Any image processing was done with Adobe PhotoShop. In a first, step we sampled down the 48-bit version to a 24-bit version of the image, which came to be dubbed the "access master." Our aim was to present legible images of the Bible pages at a reasonable file size for online remote access and for the "Gutenberg CD-ROM." The Web files were re-sized to circa 1000 X 1400 pixels, then "unsharpen mask" was applied to each one and the result was saved as a TIFF file. For the presentation, we also needed to create thumbnails and midsized images of each page. We specified PhotoShop actions for each version and applied them to the images in batch mode. The final results were saved in JPEG at a medium compression rate.

Conclusions

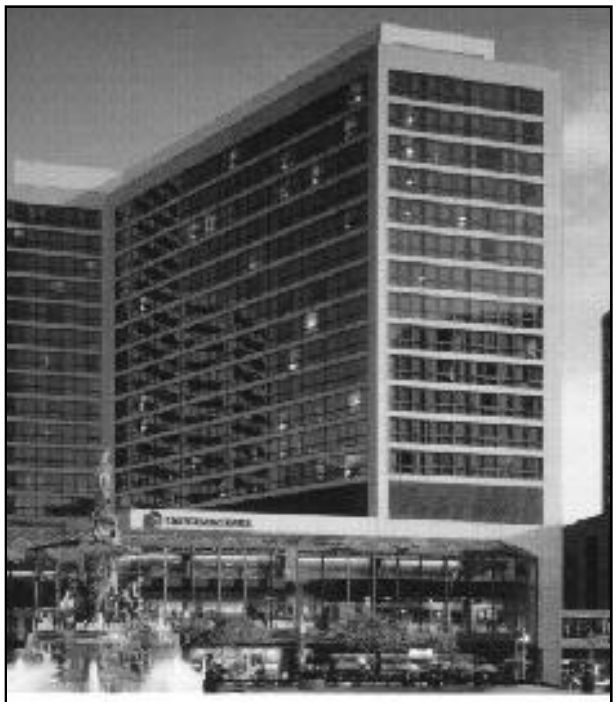
Our project to select and install a color digitization unit which meets a number of relevant conservation concerns took about one year. Minimizing the risk to the original work was our overall goal, and on reflection, we are quite satisfied with our system. For some of the components, namely the light intensity, we can tolerate the current values, but would appreciate further improvements. Recognizing that every decision in this fast-paced world of digital photography needs to be re-evaluated at regular intervals, we will try to optimize our lighting system in the future. On the other hand, we are convinced that the high resolutions our PictureGate 8000 scan back provides will be sufficient for our conversion activities for a long time to come. We can strongly recommend the digitization desk of Manfred Mayer for all kinds of reformatting activities—it is a high-quality product developed by an insider of the library/restoration community.

Is it worthwhile to undertake all these efforts? We are convinced that conservation concerns and issues in digital imaging can be worked out cooperatively between curators, preservation librarians and technicians. It should be added here that this cooperation within single institutions has to be extended to a cooperation on an international level. Direct contact with professionals in the field, as well as some mailing lists proved very helpful in our decision-making process.

A final subjective note: Our efforts in digitizing the Gutenberg Bible were at least justified by the international public, which gave us very positive feedback to our "Gutenberg Digital" presentation on the Web.

Dr. Norbert Lossau
Head of Digitization Center
Göttingen State and University Library
Dr. Norbert Lossau holds a Ph.D. in Fenno-ugric and Scandinavian studies from the University of Göttingen, and graduated as a research librarian from the University of Cologne in 1996. As Project Officer for the initiative "Retrospective Digitization of Library Materials" of the Deutsche Forschungsgemeinschaft (DFG), he is one of the coordinators for the program charged with establishing a "Distributed Digital Research Library" in Germany. He has been the Head of the Digitization Center (GDZ) at Göttingen State and University Library since 1997.

Martin Liebetruth
Imaging Service Coordinator
Göttingen State and University Library
Martin Liebetruth studied Theology and Biology at Göttingen University. Since 1997, he has been working at the Digitization Center (GDZ) of the Göttingen State and University Library as an Imaging Service Coordinator. He has conducted research in high-quality color digitization, and is the designer and co-programmer for the project "Gutenberg Digital."



The
Westin

Relax and settle into the impeccable surrounding of Concinnatir's favorite hotel. Take a dip in the rooftop pool or visit the Albee Restaurant and Lounge to sample a variety of dining and cocktail options. Explore the adjacent Fountain Square, shop, dine and discover downtown. It's all in the heart of the city and footsteps from The Westin Cincinnati.

STAY AT THE CINCINNATI WESTIN DURING THE CIMI/MCN 2001 CONFERENCE

Real Life: Virtual Experiences, New Connections for Musuem Visitors
Cincinnati, OH, USA • October 24-27, 2001

Conference Group Rate: $125 single or double occupancy

Produce, Publish and Preserve:

A Holistic Approach to Digital Assets Management

By Günter Waibel

Traditionally, museums produce and manage surrogate representations of their artifacts in the form of photography. Transparency film surrogates (4x5) have been established as the defacto standard, with applications ranging from in-house documentation to public access via print publication. With the rise of digital imaging, the creation of digital resources was initially built on the established workflow of creating photographic surrogates. Museums predominantly produced digital images as derivatives of pre-existing photography, a mere afterthought to analog capture enabled by cheap flatbed scanning technology.

Digital imaging has reached a watershed point. Present-day digital camera backs boast a resolution that effectively enables them to cut out the analog middleman. At the Berkeley Art Museum, a study by photographer Ben Blackwell concluded that files created at the maximum resolution (6000 x 8000 pixels) of a Better Light Super 6 K camera back would yield images exceeding the information stored in a 4x5 transparency. The circle closes—digital images now can be used to print 4x5 transparencies, manifesting their ability to replace the analog medium.

At a high enough resolution, it seems digital images can do everything film images can, and then some. Master image files can be re-purposed for print and Web-output, serving both promotional needs and educational goals. Museums can contribute images to networked access portals like CDL, AMICO or RLG, creating unparalleled public access to collections. Virtual access to collections may even become part of an institution's preservation strategy, since researchers can conduct most of their investigations with the aid of surrogate digital representations rather than with the original artworks.

However, the increased functionality gained in the transition from analog to digital comes at the cost of new technological challenges. The promotion of digital images from disposable, short-lived teasers on a Web site to fully fledged, mission-critical institutional assets requires distinct strategies in the production, publication and preservation of the new resource. In order to economically take advantage of the possibilities already mentioned, digital assets need to be managed in a manner resembling the way museums manage their collections of original art or documentation. Since Collections Management Systems control collection organization, it seems appropriate to suggest a database providing the same services to the rapidly increasing collections of digital surrogates. A digital assets management database supports the production, publication and preservation of digital files. It tracks a surrogate from its creation as a digital master file through its presentation as an application-specific derivative to the eventual data migration and refreshing required by changing file formats, storage media or computer platforms.[1]

At the Berkeley Art Museum & Pacific Film Archive, a database with the functionality outlined above has been under development since the beginning of the year 2000. The museum's growing involvement in networked access projects, such as MOAC (Museums and the Online Archive of California[2]) and CIAO (Conceptual and Intermedia Arts Online[3]), presented us with the challenge of having to produce: (1) digital images in compliance with project specifications, and (2) the exchange file formats the project had agreed upon. The exchange formats are EAD (Encoded Archival Description)[4] markup for navigating collections, and MOA2 (Making of America 2)[5] markup for exploring objects with-

Seven Relational Databases

The Digital Assets Management Database consists of **seven relational FilemakerPro 5.0 databases**. The different databases provide information on the surrogates, ranging from the individual files and the objects they represent to the broader context of collections and their particular hierarchical organization. All input (with the exception of Collection Level information) takes place in 10 layouts of SubOjectDB: ObjectInfo (2 Layouts), Versions, Rights, Access, Describe Structure and Sequence. Three further layouts provide orientational and navigational functions: Collection Overview, All Objects Overview, SubOjects Overview. Three additional layouts in SeriesDB facilitate the data entry for the Collections Level: Describe Structure, Add Info and Preview.

The export layout for the EAD sits in the SeriesDB, while the export layout for MOA2 has its home in SubOjectDB.

The following bulleted paragraphs give a brief overview of the seven individual databases, their functionality and interrelation. I have ordered the presentation of the databases by the hierarchy level at which they capture data.

Collection Level:

CollectionDB and **SeriesDB** have a one-to-many relationship and describe the set of Series comprising a Collection. For example, a collection of conceptual art may be represented by several series such as Videos, Installations, Performances, etc., which in turn may branch out into separate (sub)series. Within SeriesDB, individual entries take their place within the hierarchy by a system of parent-child relationships. Each Series must have a parent, and may have further children. The data captured in CollectionDB and SeriesDB make up the structural backbone of EAD mark-up by providing the collection hierarchy tagged as <coX> in the final finding aid.

Object Level:

ObjectDB and **SubOjectDB** have a one-to-many relationship and describe the set of SubOjects comprising an Object. For example, an artist's book may be represented by various pages grouped into various chapters. Within ObjectDB, individual entries take their place within the hierarchy by a system of parent-child relationships. Each SubOject must have a parent, and may have further children. The data captured in ObjectDB and SubOjectDB provides (1) the EAD with item-level descriptive information, and (2) MOA2 with the exact structural description of a complex Object.

Surrogate Level:

VersionDB, **ImagingDB** and **TranscriptDB** describe the surrogate digital files created for each object and its SubOjects. VersionDB tracks settings for master files and derivatives on an object level, while ImagingDB and TranscriptDB host the information on the actual files created for each SubOject. The data captured in VersionDB facilitates the easy setup of image and transcript files for all the SubOjects of one object. The data captured in ImagingDB populates the MOA2 markup with the individual files found in the structure of a complex Object. The ImagingDB hosts the transcript or any other related textuals and will facilitate TEI markup of text documents associated with an Object.

in a collection, i.e. the various images comprising an artist's book. In the final implementation, EAD and MOA2 build on each other: the EAD outlines the hierarchical structure of collections and supports discovery down to the item-level, from where it links to a MOA2 Object. MOA2 outlines the hierarchical structure of the item itself, presenting a navigational recreation of complex objects such as diaries or books, which consist of various different sub objects such as individual pages.

For the EAD markup, we already had most of the required descriptive metadata ready for export in our Collections Management System. The capture of the extensive metadata set of MOA2 objects, however, required a new data-architecture. Two distinct sets of metadata comprise a MOA2 Object. Structural Metadata facilitates the navigation and presentation of a digital object. In essence, it tells the viewing application what relationships exist between different parts of an object, and how the object should be assembled for the end user. Administrative Metadata facilitates the management and regulates the use of digital assets. It breaks down into technical metadata (detailing the specifications of the digital files), intellectual property rights information and metadata on the source for the digitization. All things told, a minimal MOA2 Object consists of 20 required metadata elements.

To visualize the interplay between EAD and MOA2 markup, let's consider the following examples.[6] The first block of code (Example 1) comes from a (fictional) EAD finding-aid and represents the item-level of a container list. The <daoloc> (Digital Archival Object Location) tags specify the URL for both the in-line thumbnail representation (role =“thumbnail”) of the object, and the destination of the link from the thumbnail (role =“hi-res”). However, instead of specifying a high-resolution access file, the link actually calls on the object viewer for the MOA2 xml Object.

Example 1: EAD Container List, Item Level

```
<c03 id=“bampfa” level=“item”>
    <did>
    <daogrp>
        <daoloc
        href=“http://www.bampfa.berkeley.edu/moa2/servlet
        /archobj?DOCC
        HOICE=1992.4.91.moa2.xml” role=“hi-res”><
        /daoloc><daoloc
        href=“http://www.bampfa.berkeley.edu/docs/images
        /bampfa_1992.4.
        91_136_3.jpg” role=“thumbnail”>
```

```
        </daoloc>
        </daogrp>
        <origination><persname>Theresa Hak Kyung Cha<lb>
        </pers name></origination>
        <unittitle>White Dust from Mongolia<lb></unittitle>
        <unitdate>1980<lb></unitdate>
        <repository>The Berkeley Art Museum / Pacific Film
         Archive<lb></repository>
        <unitid>1992.4.91<lb></unitid>
    </did>
</c03>
```

The MOA2 markup for one object (presented with the aid of a Java Servlet) consists of three distinct sections: a file group (<FileGrp>; Example 2), a structural map (<StructMap>; Example 3) and a section on administrative metadata (<AdminMD>; Example 4). The structural map contains the blueprint for the object's navigation, and calls on the file group to resolve the actual location of the individual image or text files. For each file (or for each specific version of the file, i.e. all files tagged "reference"), the administrative metadata grouping holds metadata on source, file-format, resolution, etc. The following examples of a (fictional) MOA2 Object illustrate how the sections interrelate to establish the complete object.

Example 2: File Group

```
<FileGrp VERSDATE='1998-09-01' ADMID='HIREZJPG'>
<File ID='bampfa_1992.4.91_136_3' MIMETYPE='image/jpg'
SEQUENCE='1' CREATED='1998-09-01'
<FLocat LOCTYPE='URL'>http://www.bampfa.berkeley.edu/im-
ages/bampfa_1992.4.91_136_3.jpg</FLocat>
  </File>
</FileGrp>
```

[repeat same structure for other versions of the same masterfile, e.g. **ADMID='LOREZJPG'**>]

Note that the markup has as many <FileGrp>s as there are versions, for example, three filegroups for three representations of the same object by high-resolution jpgs, low resolution jpgs and thumbnail gifs.

Example 3: Structural Map

```
<StructMap>
<div N='1' TYPE='paper' LABEL='page 1'>
<fptrFILEID='bampfa_1992.4.91_136_3'MIMETYPE='image/jpeg
' />
<fptrFILEID='bampfa_1992.4.91_136_4'MIMETYPE='image/jpeg
' />
</div>
</StructMap>
```

This fictional object, for brevity's sake, consists of only one image file; a complex object repeats the markup sections outlined above, and represents the hierarchy in the structural map through a nesting of tags. The N attribute on the counts of the sub objects on the current level of the hierarchy.

Example 4: Administrative Metadata

```
<AdminMD ID='HIREZJPG'>
        <FileMgmt>
                <Image>
                        <Compression>JPEG</Compression>
                        <Dimensions X='1536' Y='1024' />
                        <BitDepth BITS='16' />
                        <ColorSpace>RGB</ColorSpace>
                        <Resolution>600 PPI</Resolution>
                </Image>
        </FileMgmt>
</AdminMD>
```

[Repeat same structure for '**LOREZJPG**' or any other versions, including transcriptions.]

In order to capture all the data for the file exchange formats, we envisioned a generic database that would track all archival digital assets of BAM/PFA and allow push-button export of both EAD and MOA2 xml markup. The database would harvest descriptive metadata from our existing Collection Management System, but could also accommodate digital objects without representation in our accessioned collection, such as digital video interviews with artists in residence. Fortunately, a precedent to our endeavor existed on the Berkeley campus in the form of a Microsoft Access database used by the Bancroft Library to manage digital projects, which kept us from having to entirely reinvent the wheel.[7] The main difference between the BAM/PFA database and the existing library solution: we wanted to develop a platform-independent database we could give out to all project partners regardless of their denomination (Mac or PC), and we wanted a database which could manage all our digital assets beyond the boundaries of specific projects. Anticipating a rough ride to the realization of our vision, we named the FilemakerPro database under development DAMD,[8] or Digital Assets Management Database. A beta-release of DAMD has been available to MOAC project partners since May 2000.

Functioning as a production tool, the database establishes communication between the project administrator and the digital imaging studio. The administrator describes the hierarchy of a collection and the physical structure of the objects (i.e. a sketchbook with eight pages) in the collection, details the specifications for the master image and its derivatives, and requests transcription, if appropriate. In effect, the administrator creates a work-order, while at the same time providing the bulk of the metadata associated with the object. From here on, the production chain may be split into separate subtasks. For example, a photographer produces the master image; an intern may produce the derivatives and the transcription, as well as back up the files on a server and on the specified permanent storage medium (i.e. CD-ROM). Password access reduces the risk of accidental data corruption; each member of the team has read-only access to the database as a whole, but can write to the layouts specific to their task. A system of sign-offs on the specific tasks keeps the administrator informed on the progress, and helps co-ordinate the efforts among the team.

Once digitization for a collection has been completed, the administrator exports the metadata into the appropriate markup format. If the database gathers sufficient metadata, export-routines to different codes may enable the contribution of the same collection to various networked access projects (CDL, RLG, AMICO) in their respective native file exchange formats. So far, DAMD exports into

EAD finding aids and MOA2 objects according to MOAC specifications. Push-button export into the EAD obliterates the tedious, time-consuming and error-prone cut-and-paste patchwork of the past. The FilemakerPro database gathers and orders the different levels of a collection, populates them with the appropriate objects at the item-level, and, with the aid of a free Filemaker plug-in, writes the end result to a text file.

A further benefit of managing digital assets in a database with Web-publishing capabilities such as Filemaker consists in the option of turning the database itself into an access tool. While the creation of standardized exchange formats such as the EAD will continue to be a prerequisite for participating in networked access projects, publishing such documents to the Web requires sophisticated xml search-engine capabilities currently beyond the reach of most smaller institutions. In other words: It may be beyond your institution to provide access to collections via the exchange format demanded by consortias, but you may still contribute and then deploy your images independently by using the database itself as a local back-end.

While the production and publication of digital assets presents the community with interesting issues in the short run, another formidable challenge consists in learning how to properly care for digital files. Under less than ideal storage conditions, a 4x5 transparency has a lifespan of about five to 10 years, and since its introduction in the late 1930s, archivists all over the world have learned how to preserve film, extending its lifespan considerably. In the digital arena, such comforting knowledge does not exist as of yet. Since the lifespan of a digital file depends exclusively on the unpredictable, yet inevitable, evolution of the technical environment, clear-cut assumptions about its lifespan can hardly be made. The only certainty in a digital image's long-term future consists in its need for refreshing (moving the file to a different physical storage medium) and/or migration (moving the file to a different encoding format). However, the California Digital Library assumes that by following its guidelines, an institution can conceivably "create a master that has a useful life of at least 50 years."[9] If that claim holds true, digital files boast increased functionality while retaining a similar range of longevity in comparison to transparency film.

As Howard Besser states, technical metadata as captured in a database for digital assets management provides "the first line of defense"[10] against losing access to files, and the financial investment they represent. The database tells you about the specific circumstances of a file's creation, for example the hardware and software environment used or the compression scheme applied, and its current storage location. Moreover, it ensures consistency across a large number of files by enforcing in-house standards for creating digital media. In this way, the database facilitates the maintenance crucial to the life span of digital assets by identifying the files in need of refreshing or migration and by enabling a batch-processing approach to any impending data movement or conversion.

The metadata fields in DAMD derive primarily from two sources. For descriptive metadata, the database utilizes REACH[11] elements as a data structure, which MOAC partners have agreed to express in EAD markup for final output. As far as administrative and structural metadata goes, the database implements the necessary fields for MOA2 markup as a data structure. Since the self-proclaimed goal of the CDL's "Digital Image Collection Standards" focuses on creating a resource that will last half a century, the technical metadata (as part of the administrative metadata) and the guidelines for digital capture provide a strict standard for both the creation of digital images and their documentation. Apart from technical metadata, the set of elements referred to as "administrative metadata" includes information on rights and the physical source of the digital file. A set of elements comprising the structural metadata describes all the information necessary to view and navigate the different versions of a surrogate digital object. In the case of an artist's book, structural metadata tells the MOA2 object viewer how to create a navigational table of contents, and which page to display next.

To produce fully functional EAD and MOA2 markup, the database has to capture an impressive 35 fields per object, and an additional 20 fields for each subobject (like the individual pages of an artist's book). For all those shaking their heads at the prospect of such massive data entry, let me hasten to add: if the descriptive metadata can be imported from an existing CMS, only a maximum of four fields have to be genuinely hand-entered per object, with one additional element for each subobject iteration. Another eight elements almost pick themselves from pull-down menus (no additional element for subobject iteration), and the remaining elements either stay static enough to qualify as global fields, or the database itself works the appropriate magic behind the scenes.

Let's take a closer look at some of the key layouts in the database we developed at the Berkeley Art Museum. At first run, the database

prompts the administrator to set up global defaults concerning baseline information about the institution, on and off-line file storage and copyright statements. As I have pointed out before, the defaults reduce the amount of data-entry; however, all of them may be overwritten on a case-by-case basis during object setup.

The first regular layout of the database gathers descriptive metadata. In the case of the Berkeley Art Museum, the administrator simply types in the accession number for the object to be digitized,

and the collection management system, another FilemakerPro database, instantly populates the appropriate fields in DAMD. Furthermore, the ObjectInfo layout prompts the administrator to place the item within the hierarchy of a collection, or, to speak EAD for a moment, to define its exact destination in the container list. Collections and their levels (for example, Collection: Conceptual Art; Series: Artist's Books) evolve almost as a by-product of setting up objects. If the appropriate collection or level within a collection hierarchy does not yet exist for an object, a quick visit to the Collections layout facilitates the definition of new levels or an entirely new collection. The new level or collection will be available from pull-down menus within ObjectInfo for all further items, saving the trip to the Collections area the next time around.

In the next layout, the administrator defines the versions for all surrogates from this object. A typical work-order will include a high-end uncompressed archival file created by direct capture, two compressed derivatives for online access (reference and thumbnail), and a transcript. The order also specifies which hardware and software should be used in creating the files, essentially a request for a specific in-house workstation. File-size for all versions is defined by pixels per longest dimension, a strategy which ensures that the amount of data captured per item always stays within the same range. If file-size were defined by resolution, a scan of a very small object would yield a significantly smaller file than a scan of a rather large object. For example, a 600 ppi scan of a 1x1 object (stamp) produces a file of about 1 MB, while a 600 ppi scan of a 4x5 object (transparency) produces a file of about 21 MB. Defined by longest dimension, capture of a small object yields a very similar amount of data as capture of a very large object, deviating only by the variance in height-width proportion, not by the difference in absolute size. The file size for the stamp and the transparency captured at 3000 pixels longest dimension would amount to about 26 MB and 21 MB respectively.

Skipping both the Rights and the Access layout, which usually only require a review of the pre-defined defaults, the next crucial layout allows the administrator to describe the individual parts constituting an object. In "Describe Structure," a tree-like hierarchy defines the parts of an object and their relationship to each other as parents and/or children. The break-down of the object follows both physical and intellectual entities. An artist's book, for example, breakdown into the physical entity of pages. Each individual page yields

a digital file. However, the artist's book might also be organized into various chapters, which do not yield files, as they are intellectual concepts grouping a number of physical entities together. Since the intellectual structure provides an indispensable navigational function, it needs to be captured along with the physical entities. In the database itself, both the physical and intellectual entities define the object's structure within the same hierarchy, to be viewed in its entirety in the "Sequence" layout. The differentiation between a physical entity and an intellectual entity only requires a mouse-click. For all physical entities, the administrator requests imaging and transcription if appropriate, while intellectual concepts have no image equivalent.[12]

As already pointed out above, the "Sequence" layout displays the fruit of our labor: a hierarchical structure of the object and its subobjects as a table of contents, differentiating "children" from their "parents" by indentation. At this point, the database also analyzes the object and classifies it as "simple" (consisting of only one surrogate, like the capture of a traditional oil painting) or "complex" (consisting of more than one surrogate, like the multiple files making up an artist's book). Only complex objects have a need for the navigational functionality of MOA2, and accordingly, the database sets the EAD entity resolving the access URL for the object to either a MOA2-link for complex objects, or an image link for simple objects. Once the administrator has verified the correct setup for an object, they trigger a script generating entries into a relational database.[13] So far, we have only told the database that we eventually wish to have a specific set of versions for each sub object. The script publishes the request to the imaging and transcription database, creating entries of the individual version for each subobject. The setup for the object has been completed.

The layout "All Objects Overview" now informs the photographer that digitization for this object has been requested, and from within "Sub Objects Overview," the photographer can navigate the individual settings for each sub object and its versions. The "Visuals" layout informs the photographer of the exact nature of the request. The strategy of setting up the data before the actual digitization virtually eliminates data entry during capture - the photographer has to name the created file according to the file name generated by the database, and, at the most, type in the size of the shorter dimension. However, in our specific case, we hope to import the shorter dimension plus a host of other technical metadata directly from the cap-

ture device. Better Light camera backs allow export of metadata for each capture in a tab-delimited text file, which in turn can be imported into DAMD using the file name as a key[14]. Once the master file has been created, the photographer's assistant may produce the requested derivatives by re-sizing and compression, transcribe the object if necessary, upload the files to the server location specified in the layout "Storage," and finally, create a hard-media backup.

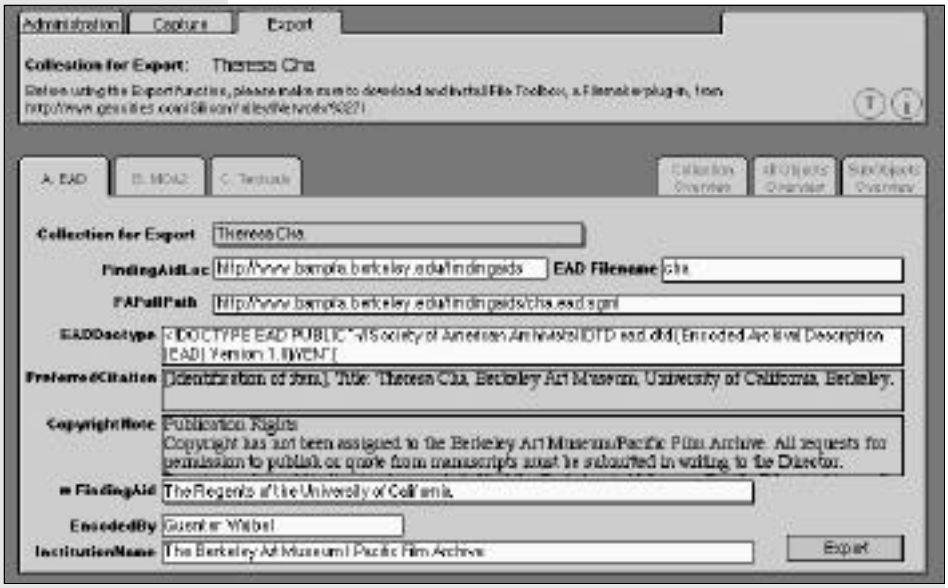
SCRIPTING

In order to keep data entry to a bare minimum, the database makes extensive use of FilemakerPro's "Scriptmaker" to automate the creation of values for several key fields. For example, in both SeriesDB and SubObjectDB, a parent-child relationship between records defines the structure of a collection or object. To resolve the parent-child relationship into a sequential numbering system which counts all sub series or all SubOjects in the order of their appearance in the hierarchy, the database triggers a two-page nested looping script. Without the script, a display of a collection or object hierarchy within one portal would not be possible. The script drills down the parent-child relationship, numbering all records along the way, until it finds no more children. At that point, it starts counting siblings, while always looking out for further children related to one of the siblings. When all siblings and their children (and children's children, etc.) of a level are counted, the script backs up to the last record with uncounted children and resumes the combined drilling for children and the lateral quest for siblings.

Another key-script saves data entry time by allowing the user to store the request for versions (master, derivative and transcription files) on an object-level. After all sub objects have been defined, the script publishes versions for each sub object to the imaging and transcription database. The script loops through all the sub objects of an object, determines whether imaging and/or transcription have been requested, and, if yes, writes the appropriate number of version files to the relational imaging and/or transcription database. In essence, each entry for a sub-object begets the number of versions requested. For a sketchbook consisting of three pages and a request for a master, access and thumbnail file plus transcription, the script creates nine entries in the relational imaging database and three entries in the relational transcription database.

The most intricate, yet ultimately most rewarding script of the database exports the interchange file format (EAD) to a text file. During data entry, the database prepares discreet parcels of EAD markup for all collection levels and all objects. Utilizing a free FilemakerPro plug-in called ToolBox, these parcels of code can be written to an external text file. The script's main responsibility lies in discovering the chunks of code in the correct order. First, the script picks up the EAD front matter and all descriptive information for the umbrella collection from a calculation field in SeriesDB. Utilizing sequential numbering for the collection and object hierarchy, the script then populates the EAD container list by working its way down the collection hierarchy levels within the SeriesDB. At each level, a sub-routine scans for parcels of item-level markup in the Sub-ObjectDB, and inserts them into the text file. Anybody who has tried to automate EAD markup knows that one of the main headaches consists in getting the different component levels (<c0x>) to close out correctly. In DAMD, the script responsible for maintaining the sequential numbering of the collection levels also features a sub-routine which memorizes all open component tags, and writes the close-out tags to the markup parcels as appropriate. The script exporting MOA2 markup operates in the same way, only one level deeper down in the hierarchy (within the object instead of within the collection).

The database now knows everything it ever wanted to know about the object and its place within a collection. After the entire collection has been populated with the appropriate objects, the remaining feat consists in having DAMD regurgitate the information in the structured fashion of an EAD finding-aid and MOA2 markup. On the EAD export screen, the administrator picks a collection from a pull-down menu, chooses a file name and (optionally) provides a preferred citation (<prefercite>) and a general copyright note (<userestrict>) for the front matter of the finding aid. The export script does the rest. It works its way down the hierarchy of collection-levels, populating them with the associated objects. Once the script has identified the record for a collection-level or item, it picks up the mark-up for the section as a complete parcel from a calculation field within the record, and adds the chunk of code to an external text file.[15]



In this way, the database assembles a complete finding-aid parsing at 100 percent EAD compliance ready for contribution to a union database without anybody even having to look at a single tag. Using a database such as the one I have described above, institutions with very little or no in-house expertise in the EAD could still contribute to EAD-based consortias. DAMD also exports the MOA2 objects at the push of a button, and soon will also feature that functionality for associated textuals. Since all the data resides within the same set of relational FilemakerPro databases, a single export-routine per collection could conceivably create the EAD finding-aid, the individual files of MOA2 markup for each object called on by the finding-aid, and the associated Text Encoding Initiative (TEI) transcription for each object.

While DAMD has not reached full maturity yet, the project already demonstrates the power of a database solution for digital-asset management. Various strategies, such as importing descriptive metadata from the CMS, importing technical metadata from the capture device, and the extensive use of global defaults and pull-down menus keep the actual number of fields for data entry at a very reasonable level. Furthermore, the database can compress the production and the publishing of digital resources economically into one single effort of data entry. Establishing communication between the various parties involved in production, and keeping track of progress, the database not only manages data, but also facilitates the management of the entire workflow. Once the metadata and the images have been captured, any number of export routines may provide access to the resource in any desired markup format. The database itself may be used as a back-end for Web access. And last but not least, the thorough documentation of your digital assets moves them into the future—it ensures that you will not have to start over again next year.

Günter Waibel
Digital Media Developer
UC Berkeley Art Museum & Pacific Film Archive
Günter Waibel (guenter@uclink4.berkeley.edu) received his education in the humanities at the University of Tuebingen, Germany, and at Georgetown University, Washington D.C. He has worked as a Webmaster at the Oakland Museum of California. Currently, Mr. Waibel holds the position of the Digital Media Developer at the UC Berkeley Art Museum & Pacific Film Archive. His responsibilities include the production, management and preservation of BAM/PFA's digital assets ranging from digital still images to digital video. He facilitates the museum's participation in networked access projects such as MOAC (Museum and the Online Archive of California) and CIAO (Conceptual and Intermedia Arts Online).

[1] Cf. Howard Besser, Digital Longevity, http://www.gseis.ucla.edu/-howard/Papers/sfs-longevity.html
[2] see http://www.bampfa.berkeley.edu/moac
[3] see http:///www.bampfa.berkeley.edu/ciao
[4] for information on the EAD, see the official Web site at http://lcweb.loc.gov/ead/
[5] for all things MOA2, see http://sunsite.berkeley.edu/moa2
[6] All examples represent valid chunks of markup; please keep in mind, however, that for brevity's sake, they only contain a fairly minimal subset of the tags available.
[7] I'd like to thank Jerome McDonough for introducing me to the more subtle details of MOA2, and Mary Elings for her introduction to the Bancroft Access Database.
[8] For the record, Richard Rinehart deserves credit for the acronym.
[9] California Digital Library, Digital Image Collection Standards. Available for download as a pdf file at http://www.ucop.edu/irc/cdl/tasw/Current/curren.html
[10] Cf. http://www.gseis.ucla.edu/-howard/Metadata/UC-May00/ppframe.htm
[11] for information on the REACH (Record Export for Access to Cultural Heritage) element set, go to http://www.rlg.org/reach.elements.html
[12] MOA2 markup itself represents the difference between entities equaling files and entities equaling concepts by relegating them to different sections of the xml dtd; only entries in the object hierarchy which have an image equivalent become part of the "File Inventory," while all structural elements (both physical and intellectual) become part of the "Structural Map." In our example, the individual pages of the artist's book could be found both in the File Inventory and the Structural Map, while "Chapter 1" would only sit in the structural Map, and from there call on the appropriate page in the File Inventory (if Chapter 1 starts on page 3, it would call on page 3).
[13] A big heartfelt thanks to Anne Verrinder, who volunteered to write the script.
[14] Dan Johnston, Senior Photographer on UC Berkeley Campus, uses a freeware tool called iView Multimedia (http://www.iview-multimedia.com), which writes the images metadata to an Excel spreadsheet, from there it can be imported into a database.
[15] A free plug-in called "Tool-Box" (available for download at http://www.geocities.com/SiliconValley/Network/9327/plPLUGs.html for both Mac and PC) facilitates the writing to an external text file from within a Filemaker script.

MCN MEMBERSHIP BENEFITS

MCN IS A NETWORK OF PROFESSIONALS SHARING THEIR EXPERIENCES WITH AUTOMATION IN MUSEUMS. AS A MEMBER, YOU WILL RECEIVE MANY BENEFITS AND NETWORKING OPPORTUNITIES. SOME BENEFITS INCLUDE:

• Receive MCN *Spectra*, which is filled with the latest information about museum automation and the activities of MCN.
• Participate in Special Interest Groups, including Topical and Regional SIGs.
• Attend the Annual Conference and workshops at a reduced rate.
• Be made aware of all publications, including the Vendor and Consultant List, and the future Topical Information Packets, and can purchase these publications at a lower rate than non-members.
• MCN makes informational mailings to its members both in conjunction with *Spectra* and separately on a number of occasions each year.

TYPES OF MEMBERSHIP

Individual Membership - $60/year
Any person interested in the goals and objectives of the Museum Computer Network is eligible for Individual Membership. They are required to pay dues as established by the MCN Board of Directors. Each person shall have all the privileges of membership including those of making motions, voting and holding office.
• Subscriptions to the Spectra, the Quarterly Journal of the MCN
• Complimentary Membership Directory
• Full access to the MCN web site
• Access and subscription to both the web-based MCN Forums and the web-based forums of the MCN Special Interest Groups (SIG)
• Annual Conference registration discount Voting privileges

Institutional Membership - $200/year
Any not-for-profit museum, library, educational institution or organization is eligible for Institutional Membership. It is required to pay dues as established by the MCN Board of Directors. Each institution shall be entitled to two representatives, who shall have all the privileges of membership including those of making motions, voting and holding office. Additional representatives from each institution are entitled to attend MCN functions at members rates.
• Two named representatives
• Two subscriptions to the Spectra, the Quarterly journal of the MCN
• Two complimentary Membership Directories
• Full access to the MCN web site
• Access and subscription to both the web-based MCN Forums and the web-based forums of the MCN Special Interest Groups (SIG)
• Web site link in special Institutional Member section
• Annual Conference registration discount
• Voting privileges for the two named representatives

Industry/Corporate Membership - $300/year
Any company or firm is eligible for Industry/Corporate Membership. They are required to pay dues as established by the MCN Board of Directors and will have all the privileges of membership except those of making motions, voting and holding office. No more than two representatives from each company/firm is entitled to attend an MCN function at member rates.
• Two subscriptions to the Spectra, the Quarterly journal of the MCN
• Discount on rental of MCN membership list
• Two complimentary Membership Directories
• Full access to the MCN web site
• Access and subscription to both the web-based MCN Forums and the web-based forums of the MCN Special Interest Groups (SIG)
• Web site link in special Industry/Corporate Member section
• Substantial discount on exhibit space
• Annual Conference registration discount

TO RECEIVE ADDITIONAL MEMBERSHIP INFORMATION, TO JOIN MCN OR TO RENEW YOUR MEMBERSHIP DUES, PLEASE CONTACT MCN AT 877/626-3800 OR MEMBERSHIP@MCN.EDU.

The Albert M. Greenfield Digital Imaging Center for Collections Academy of Natural Sciences, Philadelphia

By Eileen C. Mathias, Coordinator

The Albert M. Greenfield Digital Imaging Center for Collections germinated in the mind of an ichthyologist on staff about three years ago. A committee was formed, consisting of museum curators, collection managers and a representative from the library, to write a proposal for funding for equipment and a part-time employee to establish a digital imaging lab. Today, the Greenfield Imaging Center is a reality and is available to Biodiversity and Library staffs for digitizing images of museum specimens, as well as printed and archival materials at the Academy of Natural Sciences in Philadelphia.

The Academy of Natural Sciences was founded in 1812 by a small group of scientists, for the purpose of collecting and studying specimens from the American continent. Before that time, specimens were sent to London to be studied, classified and stored. Today, the Academy has world-renowned collections in several areas, including the largest grasshopper collection in the world and the second largest malacology[1] collection; a notable natural history library; and an environmental resources department that studies water quality and the science of limnology.[2]

The Greenfield Imaging Center is used for a variety of purposes. The lab includes one Mac G4 and three G3 workstations, each with different capabilities to accommodate a wide range of projects. The computers are equipped with Radius PressView monitors that can be calibrated. The workstations are connected to a Windows NT server and an Epson 3000 printer via TCP/IP. The four workstations include:

- A Phase One PowerPhase scanning back and Hasselblad camera for books
- An OPAL ultra flatbed scanner for single-sheet items (photographs, letters)
- A Kodak DCS 460 for specimens, and a LaserGraphics Film Recorder (to create slides)
- An APS CD-RW burner
- Polaroid Digital Microscopic Camera

Figure 1: This Blue and Gold Macaw, captured by the Phase One camera originally in color, is from Edward Lear's Illustrations of the family of Psittacidae, or Parrots, 1832. The book is owned by The Academy of Natural Sciences.

The Phase One PowerPhase workstation is used primarily for capturing images from the library's books, which include a number of beautifully illustrated folio volumes. A large copy stand, made by Tarsia Technology Industries (TTI), sits on the floor and accommodates oversized books. An adjustable book cradle is used to protect smaller items, and foam cradles are used for larger volumes. A transmissive light source, made by TTI, allows us to digitally capture images from 35 mm slides and larger transparencies. A Phase One PowerPhase scanning back on a medium-format Hasselblad camera, with a resolution of 7072 X 7072 pixels, is used for this work. This camera was chosen because of its ability to capture colors accurately. The camera captures large amounts of data, up to 144 MB at 24-bit color depth or 288 MB at 42-bit color depth. Currently, we capture at 24-bit because of time and storage restraints (four to five images per CD, if you do the math). Our hope is that these images will provide adequate detail for scholars and scientists well into the future. All digitized images are burned to Kodak Gold CDs as TIFF files, with directory structures for both Mac and PC platforms. We fully anticipate the need to migrate these images from CD to the next technology in a few years.

Color accuracy is of prime concern at the Phase One workstation, since many of the illustrated books include hand-colored plates. Book scholars are often interested in variations in color between one copy of a book and the next. In addition, color is of course important in identifying species. Kodak color strips are included in each image to provide a guide to color accuracy. Freestanding quartz halogen lights provide even lighting. The Hasselblad lens is equipped with a filter that effectively removes the colorcast created by the lights.

Color management is used to an extent in creating images. For example, the Phase One PowerPhase camera has an ICC (International Consortium of Color) profile that works quite well in maintaining color accuracy throughout the production process. Customized profiles can be created for each piece of equipment, in-

cluding scanners and printers, using specialized software and equipment such as the spectrometer. Kodak has its own Color Management System. Other products include ColorBlind, X-Rite Color, and ScanOpen by Heidelberg.

At present, ICC profiles are not embedded in each image, but the ICC profile of the PowerPhase camera is included on each CD, along with instructions on how to use the profile to convert the images in Adobe PhotoShop. At least one institution we interviewed in the process of making color-management decisions captures two versions of each image, one with and one without an embedded ICC profile, to provide flexibility in anticipation of future technologies. We have chosen not to do this because of time constraints, but if an organization has the resources, it may be wise. So far, supplying the ICC profile has worked in converting files as needed for print and Web production.

Figure 2

As stated earlier, color management is used in the imaging center, and I will explain how it is done here. Keep in mind that every lab will use color management differently, depending on products, workflow and equipment specific to that lab. Since the Greenfield lab is equipped with Apple computers, ColorSync, Apple's product, is used for color management. In the ColorSync control panel, the computer is set to use the Radius PressView calibration as the System Profile, for viewing purposes (see Figure 2). PressView monitors are compatible with ColorMatch RGB, so for production purposes the images are converted to ColorMatch RGB. This may change in the future if we migrate to other monitors, or if other RGB color gamuts, such as Adobe's, become more popular or standardized.

The scanning software that comes with the Phase One camera will not rotate images. Therefore, because of lighting issues, an image is sometimes captured and then opened in Adobe PhotoShop 5.5 to rotate that image before it is burned to CD. In such instances, PhotoShop's color-management settings are set to neutralize any chance of converting the image to another color gamut, since it is essential to retain all the information captured by the original scanning device for the archived version of the image. Once an image is converted to another color gamut, some of the data is clipped, or lost. To keep PhotoShop from converting an image, we do the following before opening the image:

- In PhotoShop, go to the File menu ▸ Color Settings ▸ Profile Setup. Set the RGB (Red Green Blue) "Assumed Profiles" to "None," and "Profile Mismatch Handling" to "Ignore." According to an Adobe PhotoShop technician this prevents PhotoShop from converting files (see Figure 3).
- Also, make sure that no boxes are checked in the "Embed Profiles" section, so that profiles are not inadvertently added to the image.

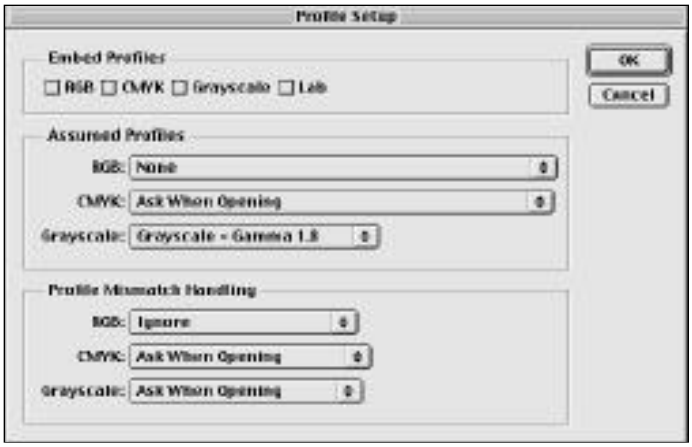


Figure 3

Once an image is archived to CD, that image cannot be changed in any way. It is fixed on the CD. Therefore, we frequently work from the archived images already burned to CD when converting images for Web use or publication. In the next example, I explain how we convert images from a Phase One PowerPhase camera's ICC profile to an RGB color space for use on a Web site. Before opening the image in Adobe PhotoShop:

- Go to PhotoShop's File menu ▸ Color Settings ▸ RGB Setup.
- Set the RGB to the ColorMatch color space, since ColorMatch RGB is compatible with the PressView monitors. sRGB is another excellent choice in this situation, since this color space was created specifically to facilititate color-safe publishing on Web sites (Figure 4). For further information about color accuracy on the web, see Trudy Levy's article in this volume of *Spectra*.
- The monitor setting, also in the RGB Setup box, is set to "Radius ColorMatch RGB," which is based on the calibrated profile of the PressView monitor at that workstation. "Display using monitor compensation" and "Preview" are checked to enable these features. This allows the user to view the image's colors as accurately as possible, albeit through the monitor's limited color gamut. It is important to note that the colors are not changed to the monitor's gamut, just viewed using the monitor's color gamut.

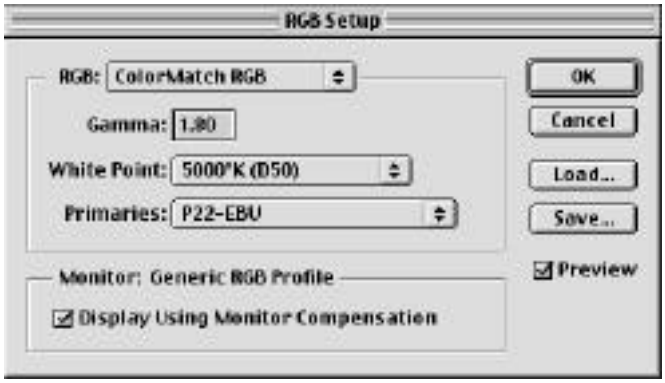


Figure 4: When ColorMatch RGB is chosen as the RGB color space, the values shown here are the PhotoShop defaults. PressView monitors use a setting of Gamma 1.8, as well as a white point of 5000K (Kelvin). These values can be set differently depending on the monitors used. Gamma 2.2 is used by PCs, Apple computers use 1.8. The 5000K refers to the luminosity of the CRT, and is the standard viewing white point for graphic arts. 6500K, a brighter luminosity, is often used as well.

• Go to Adobe PhotoShop's File menu ▸ Color Settings ▸ Profile Setup.
• Check the RGB checkbox in "Embed Profile" so that the chosen profile will be embedded into the image (Figure 5).
• Set the RGB "Assumed Profile" to the ICC profile used when scanning the image. For example, the generic profile for the Phase One PowerPhase camera gives very accurate results. That profile is copied to each of the computers in the imaging lab and placed in the Systems Folder ▸ ColorSync Profiles folder. That profile then becomes available to software applications that use color management.
• Set the "Profile Mismatch Handling" to "Convert to RGB Color." PhotoShop will then convert that image, once opened, to the RGB color space specified in the "RGB Setup" dialog box (in our case, ColorMatch RGB).

Alternatively:
• Check the RGB checkbox in "Embed Profile" so that the chosen profile will be embedded into the image.
• Open the image with "None" and "Ignore" set in the "Profile Setup" dialog box.
• Then go to Image ▸ Mode ▸ Profile to Profile.
• Set "Convert From:" to the ICC profile of the scanning device used to capture the image.
• Set "To:" to ColorMatch RGB, using PhotoShop's built-in color-management engine, or Apple's ColorSync, if you choose. PhotoShop's built-in engine works well.
• "Intent" is almost always set to "Perceptual," as this is the setting that works best with finely textured materials, such as photographs and colored illustrations with gradients.

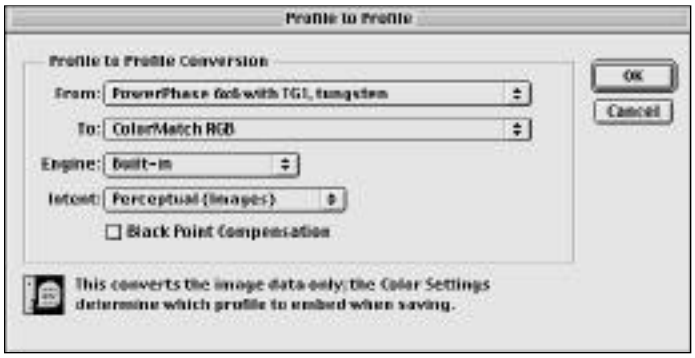


figure 5

• Do not check "Black Point Compensation." Most software applications do not yet handle this feature, and results are unpredictable.

Most printers use CMYK (Cyan Magenta Yellow Black) colors. The Epson 3000 Printer in the imaging lab actually converts images from RGB to CMYK "on the fly," so images printed from the Epson do not require further conversion from an RGB gamut. Images sent to the academy's graphics department are converted to ColorMatch RGB. The CMYK color gamut is actually smaller than the RGB color gamut, so any images to be published by an outside publisher are sent as RGB files, unless the print professionals specify otherwise. A printer will sometimes provide an ICC profile for its own printing equipment, in which case that profile can be used to convert an archived image to CMYK. One of the larger and more generic CMYK profiles is sometimes referred to as the "SWOP negative FullKey," or UCR370 of the ICC code. This can be a forgiving choice if no other options are specified.

The second workstation is equipped with an OPAL ultra-flatbed scanner, manufactured by Linotype-Hell, now known as Heidelberg. This scanner has excellent color accuracy and can attain very high, non-interpolated resolutions. The images can be saved as LAB TIFF files, a color gamut that is larger than other RGB gamuts. The flatbed scanner is used for manuscript items such as letters, diaries, photographs and 35 mm slides. The resolution is high enough that the slides are captured quite well, of better quality than images captured by many dedicated slide scanners. The Botany Department has carefully scanned their Lewis and Clark botanical specimens, attached to 11 inch x 17 inch paper, with good results. (The academy curates a collection of 226 botanical specimens obtained by Lewis and Clark on their epic journey west nearly 200 years ago.)

At a third workstation is a Bencher copy stand with two DynaLite strobe lights. Mounted on the stand is a Kodak DCS460, an instant-capture digital camera system, on a Nikon camera. The scientists are not as concerned about color accuracy as they are about overexposure of their specimens to light and heat. The Kodak DCS460 captures images in a single exposure, resulting in TIFF files of 18 MB. When equipped with bellows and various lenses, images of mollusks as small as 1 mm in length can be captured in astonishing detail. The ability to magnify is a wonderful benefit of the technology that aids our scientists in their research, and greatly enhances their ability to share the results of their discoveries with others.

Figure 6: This mollusk shell is less than 1 mm in length. The "pebbles" on which the shell rests are actually tiny grains of very fine black sand.

Figure 7: Fossilized fish, captured by the Kodak DCS 460.

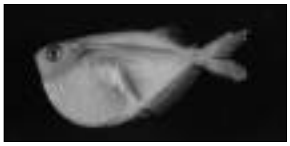


Figure 8: A wet fish on glass, with black flocking paper beneath. The flocking paper absorbs light, simultaneously removing glare and shadows.

Providing the right lighting has been a challenge, as each specimen is unique. For example, the academy's vertebrate paleontologist[3] often needs to capture images of fossils embedded in flat rock. The strobe lights are attached to a power pack that provides control over the intensity of each strobe light, independently of the other. The paleontologist sets the strobe lights asymmetrically (both in distance and intensity) so that one is brighter than the other, to create shadows, thus bringing out detail through relief of the shallow ridges and valleys created by a specimen. This lighting technique gives flat rocks a more three-dimensional look and feel, and makes it easier for the eye to sort out important details. Birds, on the other hand, are captured with the strobe lights set symmetrically, each with the same intensity, to avoid shadows, which can ultimately hide details. Plumage color is emphasized. The ichthyologists[4] frequently need to capture images of wet fish (the specimens are stored in liquid preservatives), so the fish are placed in rectangular glass containers. A shield is then placed around the camera lens, made of black flocking paper mounted on poster board, to block glare from reflections off the ceiling. The black flocking paper has proven to be very useful. It is non-reflective, and blocks glare effectively. Flocking paper is often placed under an object as background material. Velvet fab-

rics in various colors, such as black, maroon or dark blue, are also effective in the background, as a method of aesthetically emphasizing color contrasts.

Opaque sheets of various materials (plastic, fabric and fiberglass), available at camera shops, are useful in diffusing copy stand lighting. In one project, leaves and twigs of botanical specimens had been taped to paper for preservation. In capturing the images of these specimens, heavy shadows appeared which proved to be distracting. One sheet of plastic and one of fiberglass were placed together on a cardboard frame, which was then mounted in front of one set of lights. The other set of lights was raised high above the copy stand, with no fabric. The shadows virtually disappeared. Experimentation, resourcefulness and creativity are essential when solving lighting issues.

The Polaroid Digital Microscopic Camera, a recent acquisition, is used to capture images on microscopic slides of insect parts, algae and other small items. The Polaroid attaches to the trinocular head of a microscope, and the camera sees what the eye sees through the binoculars. The software has a full range of tools for color correction, focus, exposure, adding text labels and so on. The resulting images can be stunningly beautiful, depending on the subject.

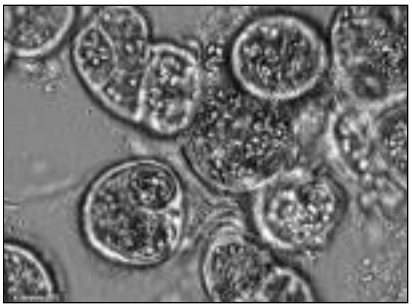


Figure 9: This algae specimen was mounted on a microscope slide over 100 years ago. By capturing it digitally, it is preserved for study by future generations.

The digital images created in the Greenfield Imaging Center have been used for a multitude of purposes. Large TIFF files are burned to CD to be archived. It is these images that are used again and again for production in other ways. Smaller JPEGs are created for Academy Web pages, often for virtual exhibitions or educational products. Adobe PhotoShop's batch-processing feature is useful in converting TIFF files to JPEGs, rotating image and other routine tasks. TIFF files are sent to the Exhibits Department to produce posters and exhibit signage, or to the Communications Department for inclusion in annual reports, event schedules and other academy publications. Several images, most often in grayscale, have been published in peer-reviewed scientific journals. Images are captured digitally when scientists from other parts of the world request a specimen. Several views of the specimen are sent via e-mail, or on disk or CD. These images often present enough detail that the scientist does not need to request the original, thus preserving it for future study. The level of magnification attained by the digital cameras has allowed the scientists to study specimens in new ways. For example, one scientist was thrilled that he can now study his newly collected fossil fishes by viewing the surrogates on his laptop while commuting on the train. By using digital images, scientists can share new discoveries with their colleagues faster and more easily than ever before, at conferences, meetings and informally.

The images are tracked internally by a customized FileMaker Pro database. FileMaker Pro is cross-platform, giving it an advantage over other readily available databases. In the future, we hope to migrate the data to an SQL database such as Oracle, for its power and scalability. The intent is to someday provide searchable access to the image database on the Web.

The greatest challenge has been compatibility of old and new. If one item is upgraded, it is often difficult to make the older equipment and/or software work with the newer. When one of the PressView monitors recently stopped working properly, it was easier and cheaper to send it out for repair than to purchase a new monitor, which would have required upgrading equipment and software at considerable expense.

The instability of the computer industry aggravates this problem of obsolescence. Six months after purchasing the Radius PressView monitors, Radius spun off the monitor portion of its business to Miro Displays, Inc. A year later, Miro Displays sold the monitor technology yet again to KDS. PressView monitors are no longer supported, that is to say, there are no software upgrades available for Mac OS 9.0, Apple's newest operating system. The PressView monitors and OS 9.0 are incompatible, and if the computers or software are upgraded, new monitors must be purchased. Short life spans of companies and quick obsolescence, as we all know, make the business of digital imaging a very expensive proposition. At this institution, we have chosen to stay "fixed in time" for the present. In a nonprofit environment, one must always balance cost versus the temptation to continually upgrade.

Because the digital imaging center is now operational, the Biodiversity and Library departments are in a better position to apply for additional funding for specific projects. The Ewell Sale Stewart Library recently received a grant from the Institute of Library and Museum Services to digitize early works affiliated with the academy. These works were some of the first published on American flora and fauna, and serve as an important historical record of the exploration of the new continent. These images will be available to the public through the library's online public access catalog.

To find out more about the Albert M. Greenfield Digital Imaging Center for Collections and for a list of planned projects, point your browser to *http://www.acnatsci.org/greenfield/*.

Eileen C. Mathias
Information Services Librarian &
Coordinator, Albert M. Greenfield Digital Imaging Center for Collections
Ewell Sale Stewart Library
The Academy of Natural Sciences

Eileen Mathias (mathias@acnatsci.org) received her BA degree from UC Berkeley ('77) and her MSLS from the University of Illinois, Urbana ('79). She worked for the Contra Costa County Public Library System for nine years in various capacities ranging from Young Adult, Children's and Reference Librarian. Then Mrs. Mathias worked as a solo librarian for an engineering consulting firm in Philadephia. Currently, she is employed at the Academy of Natural Sciences as Information Services Librarian and Coordinator of the Albert M. Greenfield Digital Imaging Center for Digital Collections.

[1] Malacology-the study of mollusks.
[2] Limnology-the study of bodies of fresh water for their biological, chemical and geological properties
[3] Vertebrate paleontology-the study of fossilized vertebrates.
[4] Ichthyology-the study of fish.

Practical Approach To Color Management

By Andrew Stawowczyk Long

Introduction

As we all know, digital imaging requires not only expensive hardware and software, but also a complex, systematic approach to achieve proper results. One of the most commonly misunderstood and difficult processes is color management. Due to a lack of user-friendly information and the high cost of hardware and software, color management still is as much an art as it is a science.

Historically, there were many attempts to ensure reliable color in the digitization process. The printing industry developed BVD/FOGRA (Bundesverband Druck—German Printers' Federation/Forschungsgesellschaft Druck e.V.—Graphic Technology Research Association), a standard that defined process colors, the color of the paper, measuring conditions and dot-gain in the printing process. Another standard was SWOP (Standard for Web Offset Printing), developed in the United States. In the realm of pre-press, the PVD (Partner vor dem Druck) system has been developed to obtain correct results. Color management was also introduced in the video and motion-pictures industry. Of all those systems, the ones still in use remain quite complex and expensive. They are also hardware and software dependent, and can not always be used with the vast number of devices and operating systems currently available.

To solve this problem, several companies formed a group called International Color Consortium (ICC), which decided to create an open color management system (CMS) that will allow users to achieve reliable color throughout the entire digitization chain. Since then, the ICC standard has been widely accepted. The standard describes color profiles that have to be created for all input, output and display devices. Other profiles (color space conversion, device linking, abstract and named color profiles) are also specified. These additional profiles are not attached to an image. They are being applied to the image data, for example, when a graphic file has to be converted from one color space to another. To make the standard more flexible, ICC created a profile connection space (PCS)—an intermediary device-independent color space which we can think of as a common translator between various devices. The ICC has chosen two PCSs—CIELab and CIExyz, both device-independent standards for viewing color developed by the Commission Internationale de l'Éclairage (CIE).

ICC profiles are files of data describing the color characteristics of a device—device settings and numeric data describing how to transform the color values from the device color space into a common color space. Most modern Color Management Systems are based on the ICC standard and use ICC color profiles.

Processes involved in color management are very complex. In this short article, I will very briefly explain CM basics, and then concentrate on describing some practical techniques.

Basic Color Theory

To better understand color management principles, let's look at the basic color theory first.

The human eye perceives color when it is being stimulated by a particular wavelength of the electromagnetic spectrum. The eye has three basic color receptors with peak sensitivity at red, green and blue. If all three basic colors are equally represented, we see gray or white. Physiological and psychological processes involved in color perception are very complex and still not well understood.

Digital capture devices and monitors work using the RGB (Red, Green and Blue) model. We call it the additive color system. Printers and some other devices use a different system called a subtractive model. They use Cyan, Magenta, Yellow and additionally blacK (CMYK) to deliver full-color images. We will explain these different models starting with an example of a computer monitor.

Imagine what happens when the computer is switched off. Obviously, the screen is black (no color), but when we switch it on and start adding red, green and blue (monitor primary colors), we eventually end up with a pure white screen (all possible colors mixed together). Contrary to this, when we use CMYK devices (let's use the printing process as an example), we start with a white sheet of paper (all colors already there). By applying Cyan, Magenta and Yellow, we subtract colors from the sheet and, eventually, we should arrive

at having a black piece of paper (no color). In reality, printing dyes are not pure enough to obtain fully saturated black. This is the reason why the fourth dye (black) has been introduced into the printing process.

Both RGB and CMYK have limited color spaces (they can describe only a limited number of colors) and are hardware dependent (colors depend on the device used and its parameters). Problems associated with being able to print exactly what we see on the screen arise from differences between the RGB and CMYK space. CMYK is a smaller color space than RGB, which means that not all colors visible on a computer monitor will necessarily print. An intelligent conversion process that transfers non-printable colors into the closest ones that can be printed has to be employed. Of course, the process is not perfect and prints always differ slightly from what we see on the screen. The number of colors displayed or printed depends also on the device used. The so-called "color gamut" describes the maximum number of colors a particular device can output.

To be able to work outside the limitations imposed by RGB and CMYK systems, CIE (Commission Internationale de l'Éclairage) developed a color space standard that is based on how the human eye perceives color. Two of these models, called CIExyz and CIE L*a*b* are the color models of choice for the ICC. Both models are device independent and describe the entire range of colors the human eye can see.

Digitizing System

An average digitizing system consists of three broad categories of devices:

Input - digital cameras, scanners, etc.
Display - CRT monitor, LCD display, etc.
Output - printers, film recorders, etc.

Short Characteristics of Input Devices

Input devices are RGB-based as they employ Red, Green and Blue filters to acquire full-color images from a color-blind CCD (Charge-Coupled Device). Most professional scanners and digital cameras are linear devices. Software for such devices allows the user to calibrate them accurately. Unfortunately, this is not the case with less expensive equipment.

Color acquired during digital capture is device dependent. To be able to recreate true colors of the captured original from a graphic file, it is necessary to acquire information about transfer properties of the capturing device. One of the most accurate and simple methods to obtain this data is by creating ICC color profiles for each device. As scanners often employ different light sources and/or different lenses for transparent and reflective scans, it is necessary to create separate profiles for each case. Similarly, digital cameras require separate profiles for each lens (differences between some lenses may be negligible) and different light sources you may employ.

Short Characteristics of Output Devices

Output devices often work in CMYK color space and the transfer characteristics of many of them, especially printers, are not linear (more about linearity later). Color is being determined by many elements of an output system. For printers they are: printing inks (dyes), receiving material (paper, plastic, etc.) and the property of a printer called "dot-gain." The amount of dot-gain differs depending on the printing surface.

Short Characteristics of Display Devices – CRT Monitors

Monitors are RGB devices. Most computer monitors are Cathode Ray Tubes (CRT). Due to its non-linearity, the monitor is the main culprit in problems associated with correct color.

Plate 1 presents the input-output graph of an ideal monitor: input intensity equals output intensity, or in other words, this monitor's characteristic is linear. Doubling the pixel value of the image (input) should produce twice as much brightness on the monitor (output). Unfortunately, the monitors in real life are different. Plate 2 shows the true characteristics of an average monitor. Fifty percent of input translates into only about 18 percent of output. This non-linearity of the CRT is called gamma and can be derived from a following equation:

$$\textbf{output intensity} = \textbf{input intensity}^{\text{gamma}}$$

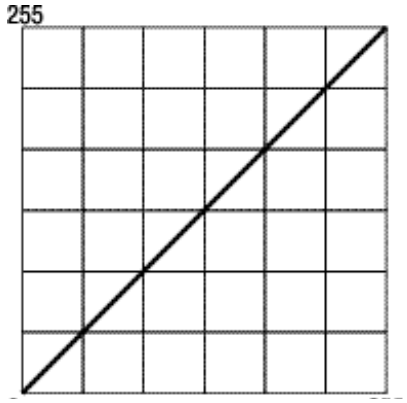


Plate 1

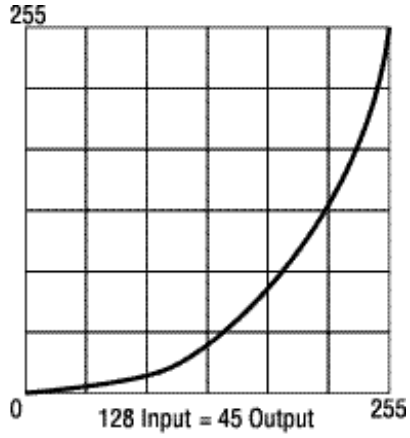


Plate 2

The gamma of an average computer monitor is between 2.4 to 2.6 (average gamma = 2.5). To compensate for gamma and display images correctly, more sophisticated computer systems (they have hardware look-up tables usually in graphic interface cards) apply an inverted gamma of 2.5 to the data (gamma 1/2.5). Monitors

compensated in this manner are said to be in gamma space 2.5. This is true for PCs, while Macintosh systems apply a partial gamma correction (gamma = 1/1.45) that results in the gamma value 1.72 on a Mac (Macintosh systems are said to be in 1.72 gamma space, often rounded up to 1.8 in popular usage). Please, make no mistake—monitors (and systems) are not calibrated when this happens, they are gamma compensated only!

Gamma is a complex and sometimes difficult concept to comprehend. In the table below, I will try to explain various gamma-related terms.

TERM	EXPLANATION
Gamma function	Gamma is the exponent of the following equation: Output = input gamma
Gamma compensation	When a gamma function of gamma = 1/2.5 is applied to an image, then the image is gamma compensated for gamma space = 2.5 viewing.
Gamma correction	When a system is linearly calibrated (gamma = 1.0), the system is called gamma corrected.
Monitor gamma	The electrical property of a CRT monitor applies a gamma function with gamma = 2.5 over an image data. To show images properly, a gamma function with an inverted gamma = 1/2.5 is applied to the image. We say that monitors are in gamma space 2.5.

Linearity and Gamma

As explained in the previous paragraph, some devices used in the digitization chain are not linear, i.e. input values do not equal output values. Systems can be set in many ways and the most common is to calibrate them to gamma = 2.2 on a PC and 1.8 (1.72 to be exact) on a Mac. These gamma spaces are, unfortunately, not the ideal environment to work in. Most times we need to manipulate images after capture by applying some adjustments (like curves or levels) and filters (like "Unsharp mask"). Gamma spaces that are not linear (gamma not equal to 1.0) create errors due to gamma compensation changes during manipulation and the linearity of tri-stimuli color spaces (RGB or CMY(k)). These errors accumulate with each operation. Some examples of the most common image-gamma induced errors include deteriorated edges, increased noise and hue shifts.

Ideally, computer systems should be calibrated to gamma space = 1.0, and all image manipulation should be performed in this space. For a number of reasons, this is usually not the case. First of all, linear gamma space is perceptually incorrect—tones from deep shadows (Dmax) to highlights (Dmin) appear to the human eye unevenly distributed. Second, often we want our systems to be compatible with other, non-calibrated computers (World Wide Web, for example), in which case gamma space 2.2 is a reasonable compromise. Of course, it is possible to work in gamma = 1.0 and then apply appropriate gamma correction over the images to make them compatible with the output, but that procedure requires extra time and effort.

Calibrating the Computer System

An uncalibrated computer system is not the ideal environment for digital imaging.

Every computer system should be calibrated before it is being used for digitisation, and on a regular basis thereafter (say, once a month). It is sometimes a tedious process and involves many steps. Ideally, the system should be calibrated using dedicated hardware and software, but it is possible to obtain very good results without them.

It is virtually impossible to describe all existing Color Management Sytems and calibration procedures. I will concentrate on just two options: one involving Adobe PhotoShop only, and a second option that utilizes a combination of Adobe PhotoShop and some calibration hardware/software, in this case X-Rite's Color Shop and Monitor Profiler (DTP92 CIE Colorimeter), a monitor color-measurement device. It seems to me that PhotoShop is the most commonly used tool by imaging professionals in the library, museum and other image archiving environments. PhotoShop's color-management system is a very powerful tool. If you do not employ any other color-management system, it should be used all the time. Do not switch it off! An additional tool supplied with PhotoShop is Adobe Gamma—a very useful (though not entirely accurate) utility for adjusting gamma and the white/black point of the monitor. Procedures described below focus on Windows-based systems, but I tried to include details relating to Macintosh computers, as well.

Calibrating using PhotoShop only—calibrating gamma, gray balance and enabling color management:

1. Turn on the computer and make sure that the monitor warms up properly (at least one hour).
2. Make sure that there is no strong ambient light present that could affect the monitor's image. Set up the normal room lighting you will use for work.
3. Make sure that your system is running in 24-bit, true-color mode.
4. Use the controls on your monitor to set it to color temperature = 6500K.

The next one is a bit tricky: We will correct for the monitor color temperature deviations (it works for monitors with fine-color adjustments only).

5. Create a white (RGB=255,255,255) large square on a black background and position it in the center of the screen. Either use a 6500K light source shining on a white sheet of paper or look at a white surface behind the window on a bright sunny day (it has to be between 11a.m. and 1p.m.). Now try to match the color of the white square on the monitor to the white sheet of paper or the white surface behind the window using the monitor color temperature fine controls. Primitive as it may seem, this technique is capable of fine-tuning the color temperature of the monitor.
6. Make sure that your PhotoShop version is at least 5.02 (there is a free upgrade from 5.0 to 5.02 available at Adobe's Web site).
7. Start Adobe Gamma by going to Control Panel and clicking Adobe Gamma icon (in Windows NT and Mac OS, you'll need to go to your "PhotoShop>Goodies>Calibration" folder).
8. Chose Wizard view. Click Next.
9. Make sure that "Adobe Monitor Settings.icm" profile is selected and click Next. If this is not the case, write down the pro-

file name, close Adobe Gamma, go to "Windows>system (system32 in Win NT) >color" and remove the profile. Start Adobe Gamma again.

10. This screen adjusts the black point of the monitor. I found it difficult to use and inaccurate. Below, I describe a much more precise procedure recommended by Timo Autiokari (*http://www.aim-dtp.net/aim/*).

Go to your Windows Desktop by dragging the Adobe Gamma window out of the way and right-click on it. Chose "Properties>Appearance." Make sure that the box "Item" shows "Desktop." Click on "Color" properties and change it to pure black. Click "OK."

Adjust the contrast control on your monitor to maximum.

Set the brightness control of the monitor to maximum. Locate the vertical height adjustment of your monitor and set it to some smaller size, so you can see a narrow black border at the top and bottom of the main image (Win desktop) area in the center. Now, the Desktop's canvas should not be quite black.

Turn the brightness down until the central, desktop area is just as black as the narrow strips at the top and the bottom. Leave contrast at maximum. The black point is set.

11. Drag Adobe Gamma window back to the center of the screen. Click Next.

12. Set the monitor phosphors to Trinitron or P22-EBU (they are used exclusively in modern computer monitors and there is not much difference between them). If you know your monitor phosphor chromacity values, you may select "Custom" and type in the values. Click Next.

13. Set the desired gamma value to 2.2 (1.8 on the Mac). This option is not available in Windows NT or on older systems, which cannot control monitors. Click "View single gamma only." Adjust the slider until the central square blends with the surrounding frame. To do it properly, move away from the monitor at least 3 feet and squint your eyes.

14. Uncheck "View single gamma only" box. Adjust red slider if required. Adjust blue slider if required. It is possible that you will need to adjust the green slider as well. Repeat above adjustments as long as required. Click Next.

15. Set the hardware point at 6500K (daylight)*. Use the measure utility if you are not sure it is accurate. Click Next.

16. Set "Adjusted White Point" to "Same as Hardware." Click Next.

17. Click Finish and Save to save profile as "Adobe Monitor Settings.icm."

18. Calibration is completed.

19. Open PhotoShop and go to "File>Color Settings>RGB Setup." Check the "Display using monitor compensation" and "Preview" boxes. Set RGB to "Adobe RGB (1998)"—it is probably the best color space for practical use. However, if you need to preserve a greater range of colors, you may choose CIE RGB or Wide Gamut RGB. (Note: since these are very large color spaces, not all colors can be displayed on a monitor, resulting in errors if you do any manipulation of images.) On a Mac, select Apple RGB. For Web-based work, the best option is sRGB, but images saved in this color space should not be used for archiving—sRGB is a limited color space.

Make sure that Gamma is 2.2 (1.8 on a Mac), the white point at 6500K and primaries "Adobe RGB (1998)."

20. Save the profile under a unique name in "Windows>system (system32 on Windows NT)>color" folder.

21. Open "File>Color Settings>Profile Setup." Make sure that all boxes in "Embed Profiles" are checked. For "Assumed Profiles," it is probably the best option to choose "None" for all. In "Profile Mismatch Handling," select "Ask When Opening."

*Some researchers (Anne R. Kenney, Oya Y. Rieger: Moving Theory into Practice (California: Research Libraries Group, 2000)) recommend 5000K. Personally, I prefer 6500K, but the choice is purely subjective.

Calibrating using PhotoShop, X-Rite's Color Shop and Monitor Profiler (Optimizer)—creating an ICC color profile for the monitor, calibrating gamma, gray balance and enabling color management:

1. First, go through the procedure described above in "Calibrating using PhotoShop only." Remember that all preliminary conditions (monitor warm-up, correct lighting, etc.) have to be observed.

2. Install Color Shop software on your machine (it comes in Windows and Macintosh flavors).

3. Plug in (USB or serial) and power up the X-Rite profiler.

4. Start Monitor Profiler program.

5. The program asks to adjust white and black point of the monitor; however, do not change contrast or brightness settings if you already calibrated the monitor with Adobe Gamma as explained above.

6. After selecting the correct device (monitor optimizer) and the right computer port (USB or COM) in the opening window, position the Profiler on a piece of black cloth, check the "Calibrate device" box and press OK. After calibrating the device, the program will ask you to attach it to the screen via a suction cup. Click "Calibrate" and wait for the program to finish the procedure. It may be advisable to cover the entire monitor and the Profiler with a dark cloth.

7. Save the profile in "Windows>system (system32 on Windows NT) >color" folder using a meaningful name (eg. "27_July_00.icm").

8. Open Adobe Gamma. Select "Step by Step Wizard." Click Next

9. Click "Load" and select the profile that you just created and saved. Click Next.

10. Go through all the steps in Adobe Gamma. If you already went through the initial calibration, you should not need to adjust any controls.

11. Save the profile.

12. Make sure in PhotoShop under "File>Color Settings>RGB Setup" that the monitor uses the new profile.

Calibrating Scanners and Digital Cameras

High-end scanners have built-in procedures for calibrating and creating ICC profiles. Even such scanners can benefit from some additional calibration.

To calibrate a scanner, most commonly we will use Kodak Q-60 Color Input Target, also known as an IT8 target (Plate 3). It can be also used to calibrate digital cameras; however, GREGTAGMACBETH ColorChecker Color Rendition Chart (available from Kodak) is also used. Calibration has to be performed separately for each different kind of material that we intend to scan. For reflective scans or digital cameras, we need to use a reflective target (Q60R1). For

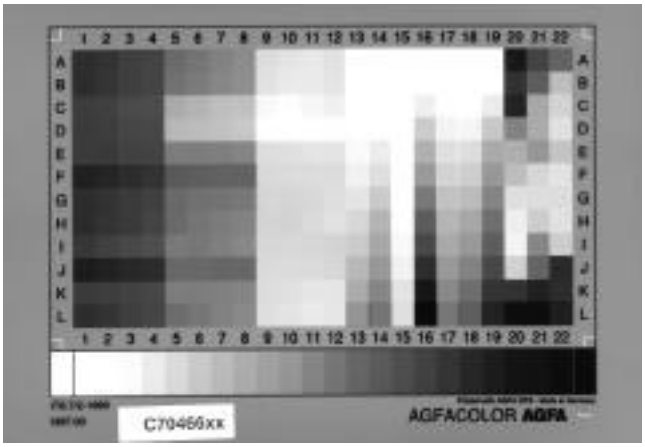


Plate 3

film, we need targets that were made for a particular type and brand of film.

As always, the best results are obtained using a specialized software that reads the scan of the IT8 target, compares it to reference data created in a lab, then creates an ICC color profile for the scanner. Examples of such software are Kodak ColorFlow ICC Input Profile Builder (www.kodak.com), WiziWYG or Compass Profile from Praxisoft (*www.praxisoft.com*), and ColorBlind MatchBox from ColorBlind (*www.color.com/Products/matchbox.html*). By paying more you get a much more sophisticated system, but even inexpensive software like WiziWYG can do a very good job.

Calibration using specialized software is very straightforward and yields excellent results. However, it is possible to calibrate a scanner (or digital camera) very accurately using just PhotoShop, an IT8 target and a digital image of a computer-generated simulated Q-60 target. Such simulated targets are available from AIM (*www.aim-dtp.net/aim/calibration/kodak_q60*). The procedure involves scanning an IT8 target and comparing it with the simulated one in PhotoShop.

Below, I describe a calibrating procedure developed by Timo Autiokari from AIM.

Note: Because simulated targets are small, they have to be re-sized to the size you will be working with. As they are just patches of color, you can re-size them to virtually any size you want. Just make sure that you use the "Nearest Neighbor" resample method, not "bi-cubic" or "bi-linear." Also, make sure that you convert them from the AIM color space into your system's color space. Another important point is to make sure that you use a simulation target that matches your real target (they all have identifying numbers).

1. Acquire a simulated target from AIM.
2. Scan Kodak IT8 target on the scanner you want to calibrate.
3. Using layers in PhotoShop, position the scan on the top of the simulated target and set a "Difference" mode for the top layer. This technique allows you to compare two images. If the colors are exactly the same, the result is black. In this way, it is possible to visually adjust the scan.
4. Use levels (or curves) to adjust white patch (Dmin—leftmost patch in the grayscale at the bottom) in the simulation target to RGB=255,255,255.
5. Create an adjustment layer (Curves) on top of the IT8 scan layer. Do not apply any adjustments yet.
6. Create an adjustment layer (Hue/Saturation) on the top of the Curves layer. Do not apply any adjustments yet.
7. Create an adjustment layer (Levels) on the top of the Hue/Saturation layer. Do not apply any adjustments yet.
8. Group all the adjustment layers to the IT8 Scan layer (position the mouse cursor on the line dividing the layers (in Layers dialogue box) and, while holding down <ALT> key, click the left-hand mouse button).
9. Using the white-point slider in Levels adjustment layer, make the Dmin patch (leftmost patch in the horizontal grayscale at the bottom) as black as possible.
10. Use Curves adjustment layer to make grayscale patches as black as possible.
11. Use Hue/Saturation adjustment layer (using Master channel) to make color patches in the IT8 target as black as possible. Use Saturation and Hue sliders, but NOT the Lightness slider.
12. Use individual channels in the Hue/Saturation to make patches as black as possible.
13. Procedures nine through 11 may need to be repeated many times until correct results are obtained.
14. Save the Curves, Levels and Hue/Saturation settings.
15. Write an action by means of which you will be able to apply the corrections to scanned images automatically. The action should include:
16. Converting into 16-bit mode (for greater accuracy).
17. Applying saved Levels adjustments.
18. Applying saved Curves adjustments.
19. Applying saved Hue/Saturation adjustments.
20. Converting back to 8-bit mode.

An improved procedure is described in detail at
http://www.aim-dtp.net/aim/ calibration/kodak_q60/calibrate2.htm
Similar procedure can be applied for calibration of digital cameras.

Calibrating Printers

To calibrate printers accurately, special software and hardware is needed. Best results are obtained with spectrophotometers—devices that can measure hundreds of color patches of a special target printed on the printer that is being calibrated. Results of those measurements are then automatically compared with a reference file, and an ICC color profile is created. Relatively inexpensive software and hardware can be obtained from X-Rite (*www.x-rite.com*).

It is possible to calibrate printers manually, but usually it requires a lot of time and printing material. Also, with many printers, you will need to repeat calibration procedure on a regular basis and every time you change printing inks, paper or chemicals (if the printer uses them). Some high-end printers (like Fujix Pictrography) have special calibrating devices that permit an ICC profile to be created just once and used accurately even if printing materials have been changed.

Manual calibration of a printer can be achieved with PhotoShop. In "File>Print>Setup>Transfer," it is possible to adjust transfer curves for each individual printing color. Using an IT8 image or a simulated IT8 target (on the calibrated computer system), you can adjust the printer by comparing the original image to the printout.

Note: It is very important to view originals and prints under correct lighting conditions. Ideally, you should invest in a professional viewing booth; however, they are quite expensive. The next best option is to buy a continuous spectrum (Color Rendering Index, (CRI) between 98 and 100—98 is bare minimum), 6500K (daylight) corrected light, and set it up in a place where there is no interference from other light sources.

Digital Imaging Workflow

Providing that we calibrated our imaging system and all input and output devices, the correct digitization workflow shall include:

1. Acquiring images (scanning, capturing using a digital camera, etc.).
2. Applying color calibration to images by means of an input device ICC color profile.
3. Editing images (to avoid errors this should be done in gamma 1.0 and on a calibrated computer system. However, most PC-based systems are calibrated to gamma = 2.2 to compromise for displaying images on uncalibrated PCs and Macs. Usually, errors introduced by gamma = 2.2 are not huge).
4. Saving images with embedded ICC color profiles.
5. Applying color calibration to images by means of an output device ICC color profile.
6. Outputting image (printing, writing to film, etc.)

Let's take a closer look at particular steps.

Step 1. Many input devices' characteristics are not linear. To correct for this and to calibrate the white point, ICC color profiles shall be created for each capture device. Some manufacturers include color profiles with their devices. We have to be aware, however, that profiles used with digital cameras and scanning backs depend on the light used during the image capture. Ideally, every user should create color profiles for each such device for all light sources that will be used during digitization.

Step 2. If a color profile has been written into the image by the capturing device, the image can be saved for archival purposes. If not, then a profile (specifically created for the capturing device) has to be applied to the image. Applications like Adobe PhotoShop allow to convert a non-profiled image, providing that we have an ICC profile for the capturing device (after opening an image, go to IMAGE > MODE > PROFILE TO PROFILE and select relevant ICC profiles conversion – FROM: input device profile that you created, TO: your system display profile).

Step 3. Ideally, editing should be done in gamma space 1.0. Only in gamma 1.0 applying filters, using curves or levels will introduce a minimum of errors. Unfortunately, most systems are calibrated to either gamma 2.2 (PCs) or 1.72 (Macs). If you choose to calibrate your system to gamma 1.0, you will find it difficult to work with some applications. It is especially annoying if you use the computer to access the Internet. The images will appear lighter than normal and flat. If your system is calibrated to a gamma other than 1.0, errors introduced during image manipulation may not be huge, but they may be visible especially in shadows. (See the paragraph on Linearity and Gamma).

Step 4. After editing images, they should be saved with the embedded ICC color profiles. PhotoShop does it by default. It will allow users to open images correctly on systems that do not have the profile for the input device. At this step, images should be written to a media for archival storage purposes.

Step 5. If we want to output images than another profile conversion is needed. This time we convert the image FROM: your system display profile TO: output device profile.

Step 6. Outputting the image.

Preservation Issues

Unfortunately, it is very difficult to find anything on this topic. It seems that preservation practitioners still have not make up their minds about color profiles—same as most other aspects of digital imaging. ICC profiles, however, are becoming de facto standards and there is a good chance that they will be recognized and supported by future systems. In the worst case scenario, they will be ignored. As they are being written into header tags (in TIFF files), they do not corrupt the image data. Colin Webb, Director of Preservation Services Branch in the National Library of Australia told me that to his knowledge, no one so far discussed those problems in detail from an archival point of view.

The following file formats support ICC color profiles: TIFF, Photo CD, PNG, JPEG, EPS, PCT, PDF, PSD.

Acknowledgment

I would like to thank Timo Autiokari (*www.aim-dtp.net/aim*) for his excellent ideas on practical digital imaging, some of which were used here.

Bibliography

Anne R. Kenney, Oya Y. Rieger, Moving Theory into Practice, Digital Imaging for Libraries and Archives, (California: Research Libraries Group, 2000).

Timo Autiokari, Accurate Image Manipulation for Desktop Publishing (*http://www.aim-dtp.net/aim/*).

Guides to Quality in Visual Resource Imaging, Research Libraries Group, (*http://www.rlg.org/visiguides/*).

Technical Guides, Adobe Systems Inc., (*http://www.adobe.com/support/techguides/photoshop/*).

ICC Specifications, International Color Consortium, (*http://www.color.org/*).

Charles Poynton, Poynton's Color FAQ, (*http://www.inforamp.net/~poynton/notes/color_and_gamma/ColorFAQ.html*).

An Introduction to Digital Scanning, (Agfa-Gevaert, 1994)

Andrew Stawowczyk Long
National Library of Australia
Andrew Stawowczyk Long (astawowc@nla.gov.au) has been a professional photographer for almost 20 years and a professional Information Technology Officer for the last seven years. He has been involved in numerous imaging projects and provided advice on digital imaging to various institutions. Currently, he is working as a Photographic Services Manager at the National Library of Australia.

Do You See What I See?

By Trudy Levy

Color informs our visual perception. Besides the original intent of the color itself, color also influences our perception of shadows and highlights, intensity and subtlety, and contrast. Without accurate color rendition, we do not fully perceive an image. Yet, we all have seen for ourselves that colors change as we view the same image on different workstations. How can we share an image electronically so others see what we see? The answer: We must manage the delivery of color. In other words, we must control the three processes which define color in electronic display —gamma, brightness and hue—at all stages of a digital image file's existence. Welcome to the world of color management.

The different stages at which color management plays a crucial part in an image's life:

- Creation stage—the scanner or digital camera and its software
- Correction stage—image workstations and image manipulation software
- Production stage—workstations and multimedia authoring software
- Viewing stage—the end-user's monitor, its display settings and viewing environment, plus the software displaying the image

Fortunately, the printing world has been trying to solve the problem of color accuracy for quite some time; with the advent of web printing, first steps have been taken toward addressing color issues in online delivery. For the near future, we will have an interim stage of evolving Web images. You may have to trade the rich color depth of which digital technology is capable for consistency of appearance when creating your Web images. However, you can still prepare for the future when master archive images will be the direct source of electronic display, while in the meantime producing Web images which are good and consistent within the current parameters. In order to be prepared for the future while making the most of the present, you must learn about color management.

Please note:

I will be discussing images meant for electronic display only. Images meant for other publication techniques or archival purposes will have other color needs. All Web sites mentioned in the body of this text are listed as complete URLs in the resource section at the end of this article.

Color Management and Calibration

Controlling the color rendition of images is called color management. As we just noted, different workstations display colors differently. The current technical description is that a monitor inhabits its own device-dependent color space. A color space is a collection of mathematical formulas (machine language) by which a device knows how to display color. The goal of a Color Management System (CMS) is to establish a device independent color space to which each device refers as a common language. The most common device independent color space is CIELab. Both ColorSync (Apple) and ICM2 (Windows) use it. There are also good proprietary systems such as ColorBlind (ITEC), which can be used to calibrate your devices as well.

Calibrating your workstation to a color space means creating a profile for each piece of equipment. Most of the time, the calibration process tends to feel rather complex, like tuning all the instruments in a symphony orchestra simultaneously. However, when all is said and done, the process is as simple as "match this to that." In fact, you may have already been requested to calibrate your monitor by an application such as PhotoShop. Or, if you use a Kodak product, you may have seen references to ICC (International Color Consortium) profiles. The ICC, a vendor organization, leads the effort in establishing standard profiles and open-platform color management systems. When you calibrate your equipment, in particular your monitor, you are creating a profile of the gamma, brightness and hue settings in a particular color space for your images. By mapping device-dependent characteristics to a device-independent color space as a common exchange language, the CMS makes profiles meaningful to devices such as your printer and other monitors viewing the same image.

Gamma

Anyone who works with digital images should know about gamma. Many people work with gamma without knowing what it actually is. They do know that gamma is adjustable and that most monitors are set at 2.2, except Apple's monitors, which are set at 1.8. They also know that when an image is created or modified on an Apple monitor, it appears darker when viewed on a non-Apple monitor. The term "gamma" comes from the Cather Ray Tube (Television) technology, to which computer monitors are related, and is an exponent describing the distribution of the intensity of light. The start-

ing point, black, and the finish, white, are not impacted by the gamma, but the rate of increase is. As a result, the apparent brightness of an image, mainly in the mid-tones, will appear differently in different gamma environments.

Timo Autiokari shows several examples of the effect of changing the gamma setting on the perception of images in his very helpful Accurate Image Manipulation Web site. However, it is important to note while reviewing Timo's work that his discussion revolves around the printing process, which has different needs than electronic display.

Brightness

Brightness is the intensity of light. It is best understood in a sales pitch for high-end LCD projectors. As the projector vendor will tell you, "the brighter the projection, the greater the range between black (no light) and white (all light), thus the greater the range for each of your hues." Furthermore, not all lights have the same color of white. We all know about the yellow cast of tungsten lighting. This difference is expressed as the white point of a light source. When you use a digital camera, you may have a "White Balance" option, which adjusts the color balance to remove the color cast the current lighting conditions may cause. You can also adjust your monitor's white point to reflect different lighting conditions. Tungsten lighting has a white point of 2800K, while daylight has 6500K and computer monitors 9300K. When you calibrate your equipment, you establish in its profile its range of brightness and the white point for the display environment.

Hue

Hue is probably self-explanatory. The additional factor in electronic display is not what hue, but how many. Actually, we can't even see all 16.7 million possible color definitions most computers now offer, but they do enable us to show greater detail and blending of colors. When I started out in this field, I was asked to put together a slide show on a machine which was only capable of displaying 16 colors. Fortunately, those days are long gone. If you want to see what difference more colors make, just change your monitor display to fewer colors. In calibrating your equipment, you establish the brightness or range in each of the three colors—Red, Green, and Blue—used in electronic display. Then you select a Color Space.

As a quick aside: you may have heard about Web-safe colors. They are "safe" in that they tend to look the same across different platforms. These are specific individual colors, however, and they do not help in the quest for displaying photographic images accurately. They are the colors you might choose for a graphic element, text or background on a Web page. If you are interested in that area of color, you might look at Lynda Weinman's site. She is a leading Web Graphics Guru. For viewing photographic images with accurate color there is no safe color, only safe color spaces in which many display devices can perform.

Color Space

As mentioned earlier, the most common color space used is CIELab. Kodak has its own version of CIELab called PhotoYcc, which is used in the PhotoCD ImagePak format. Be aware that you will lose the PhotoCD's CMS capability if you edit or modify the images within your system. For the time being, however, this is a moot point as current browsers cannot perceive these color spaces. The Web world is developing sRGB, a Standard Default Color Space for the Internet which was introduce in 1996 by Hewlett Packard and Microsoft. At this time in the Web's evolution, sRGB is the color space in which most display devices can function accurately, making it the color space of choice for displaying images on the Web. Many image software companies, including Corel and Adobe, have adopted sRGB. In fact, the default color space in PhotoShop 5.5 is sRGB, although you can choose CIELab if you wish. While sRGB has its limits, at this time, it is the best color space for electronic display. Keep in mind, however, that it is a limited space with a narrow color range and certainly not appropriate for archival purposes.

As a sidebar: PhotoShop 5.5 also defaults to a gamma of 2.2, which as mentioned earlier is the standard for most monitors. In this way, most of the time color will be consistent between display devices.

Other groups are developing color management systems that can inhabit broader and deeper color spaces in order to replicate their image's color as accurately as possible. If you work in the press world, you might have access to the International Press Telecommunications Council's Digital Newsphoto Parameter Record (DNPR) version 4. As the tech specs state: "This is the lower-level file format for encapsulating digital news photographic data. It allows for editorial and technical information to be carried in the same file. Version 4 just released incorporates datasets for an ICC Device Color Profile."

Okay. Gamma is set, the white point and brightness range in each color established and a color space chosen. Ready to manage your color for the Web? Let's start with the first stage: Creation and Correction.

Controlling the Processes at the Creation and Correction Stages

The first step is to calibrate the equipment used at the creation and correction stage. This calibration will become the profile information for your images. Some of your devices come with profiles. If you go into your monitor control panel, you will be asked to select your type of monitor. Your color management system can then plug in the factory suggested device profile. If you have newer operating systems with more fully developed color-management systems, you can also create your own custom profile for your devices. This can be especially important as ambient light and environment will have an impact on display.

The following is a series of steps that the MacOS system walks you through to calibrate your monitor. I am using the Monitor Calibration Assistant as an example because it explains the process. You can do something similar in Windows under Display/Settings/Advanced. Windows will allow you to choose sRGB color space profile, or select your monitor, set gamma and adjust color. To adjust color, you import a bitmap image, then adjust your color on your monitor until it looks like the image. On both platforms, you can then save your custom profiles.

The first screen after the overview has you adjust your brightness. You set your contrast at its highest and then set your brightness so that a shape is barely visible against the background. Note that ambient lighting has an impact on the apparent brightness of a monitor, so try to make the ambient lighting as consistent as possible.

The next step is beginning to set the gamma. In this screen, you adjust the midrange of the three color channels—Red, Blue, Green—to balance so that the Apple (object) blends with the background.

Then you select your monitor's color characteristics. Cather Ray Tube makes images by means of phosphors, which determine the range of colors you can see on your screen, as shown below:

If you want consistent Web images, you should select the sRGB profile, which will safely meet most other monitors capabilities.

Then you must select the gamma setting for this monitor. Again, for Web images, it is recommended that those of you using an Apple monitor in this non-Apple world adapt by preparing your images for viewing on a monitor with a higher gamma. You can do this by either resetting your gamma setting on your Apple monitor to 2.2 to mimic the other monitors, or by using PhotoShop to adjust the gamma level of your images to .8 in the Image - Adjust Levels dialog box. Adobe has a quick little tutorial on their site about this adjustment process.

The final step is to set the white point of color to the perceived light you are using. This is usually expressed in terms of daylight or tungsten terminology, as in different lighting situations, warmer or cooler light. D6500 is a standard term.

So now all the images which are created or manipulated using this monitor have a profile which can express digitally what the colors are that you wish to have displayed, and most monitors are capable of displaying them. How do you communicate that to the viewer's displays?

Controlling the Processes at Production and Viewing Stages

If you do not implement your Web images yourself, you can still try to maintain management of the color. For example, if you provide master archive images to an outside service provider, then include with those images the color profiles and insist that this information be embedded in the file and that they not manipulate the image, but only convert and reduce its size. If they do manipulate the image, insist that they maintain a color management system which will provide accurate color profiles to be embedded in the final Web image. Many file types can contain this information and many imaging software packages can read it. The common image file formats such as Tif and Jpeg, as well as Kodak PhotoCD, have this capability. PNG images, which are actually slated to replace Gifs by the World Wide Web Consortium, and Mr.Sid images, a highly compressed file format by LizardTech, can also contain ICC profiles. Of course, if you are producing your own Web images, then you should expect the same from yourself.

So now that we have color information, how do we relay it to the viewer's display device?

As browsers cannot yet read the color profile information, you can try some interim steps. For starters, you can simply publish the information on your site, enabling viewers to calibrate their monitors to match your specifications. An example for this strategy can be found on the Accurate Image Manipulation for Desktop Publishing Web site:

"The AIM color space (and the color space of my original images) is:

- Trinitron Primaries (these are the most common phosphors used in the CRT monitors).
- White-Point D6500 (this is the daylight, the hue of the actual sunlight at noon).
- Gamma 1.0 (the linear light domain, free from all gamma-induced errors).
- A copy of the original is then gamma compensated by 1/2.5 (inverse gamma 2.5) and uploaded to WWW. This is effective from 1.1.1998, some earlier photos are in gamma space 2.2. In addition, for some specific purposes, there are images with different file-gamma, these are explained on those pages."

Or perhaps you might intrigue someone to create an online calibration routine such as Imation and Ecolor provide at their sites. With such a program, you "help" your viewer calibrate their monitors to match your managed color system.

Ecolor and Imation are software developers providing e-commerce sites with the ability to show true color. Ecolor has you calibrate your monitor online at its Web site. Then, the site places a cookie in your drive, so when you visit a site which they service, it will tell the image server what your monitor profile is. The server then sends you images which your monitor will project accurately. At this time, the services costs approximately $1,500 a month. Soon the browser technology will catch up and be able to perceive color profile information, without needing this intermediary program.

And when the browsers can read it, you will be ready with all your color information. Plus you will no longer need to be restricted to the sRGB color space or gamma 2.2. The advent of digital imaging has brought digital color and the ability to manage it. Those

creating digital images who understand how color is produced on electronic displays can calibrate the pieces of equipment involved in creating and displaying digital images, thus creating accurate color profiles. Yes, now instead of saying the red was the color of blood, we can accurately say the red was 54, 73, 43.

For those wishing to learn more, I have assembled a list of resources.

Resources

• Other People's Guides to Web Publishing or Preparing your Images for the Web

• PhotoShop - the secrets of good gamma
http://www.adobe.com/web/tips/totalphs9/main.html

• Chapter 6: Adding Images to Your Site by Philip Greenspun, part of Philip and Alex's Guide to Web Publishing *http://www.arsdigita.com/books/panda/images.html*

• Adobe Online services will prep your images for the web
http://www.adobe.com/services/main.html

• Accurate Image Manipulation for Desktop Publishing
Okay, this isn't about Web publishing, but Timo has some great examples of "Gamma-induced errors" and a lot of other good information. *http://www.aim-dtp.net/aim/index.htm*

Examples of Online Calibration

• Ecolor - click on True Internet Color setup
http://www.ecolor.com/01/01_01.html

• Or try out Imations Verifi's calibration
http://www.imation.com/products/printing/content/0,1011,1618,00.html.

Color Management Discussions and Examples

• Color Management Workflows for PhotoShop 5.0
http://www.adobe.com/support/techguides/photoshop/cms2/cmwork.html

• David Bourgin's Color Space FAQ published on the AIM site. Complete but readable discussion of Color Space
http://www.aim-dtp.net/aim/technology/color-spaces_faq/color-space.faq.txt

• Apple's introduction to color management
http://eworld.com/colorsync/benefits/training/elements.html

• Introduction to Color Management in Windows
http://www.asia.microsoft.com/hwdev/devdes/icmwp.htm

More on Digital Color

• Understanding Digital Color By Phil Green
http://twinpentium.lcp.linst.ac.uk/digitalcolor/

• Kodak: Digital Imaging Fundamentals
http://www.kodak.com/daiHome/DLC/book3/chapter2/index.shtml

• The Ultimate Internet Coloring Database
http://www.colorpro.com/info/data/projects.html

• Browsers; Color-Management Support
http://www.dnai.com/~ccowens/browser_color/test.html

Further Information on Publishing Controls Being Developed.

• PNG—Portable Network Graphics
http://www.libpng.org/pub/png/

• IEC publishes first standard in a series on color management
http://www.iec.ch/nr1899.htm

• A Standard Default Color Space for the Internet—sRGB (historical)
http://www.w3.org/Graphics/Color/sRGB

• W3C working Draft on Color
http://www.w3.org/TR/2000/03/WD-SVG-20000303/color.html

• Not Just Decoration: quality graphics for the Web
http://www.w3.org/Conferences/WWW4/Papers/53/gq-boston.html

The International Color Consortium

The International Color Consortium was established in 1993 by eight industry vendors for the purpose of creating, promoting and encouraging the standardization and evolution of an open, vendor-neutral, cross-platform color management system architecture and components.
http://www.color.org/

International Press Telecommunications Council was established in 1965 to safeguard the telecommunications interests of the World's Press.
http://www.iptc.org/

Lynda Weinman's site, leading Web Graphics Guru
http://lynda.com/

Trudy Levy
Image Integration—The Digital Imaging Guide
Trudy Levy (Trudy@dig-mar.com) consults on converting slide libraries into digital archives and on general issues of integrating digital imaging into the work process. As the founder of "Image Integration—Digital Imaging Guide," she is a recognized expert in visual information and speaks on digital imaging to organizations such as the *Visual Resource Association*, the *Special Library Association* and the *Society of Marketing Professionals*. She has written on managing visual assets in industry journals such as *InformationWeek* and *MacWeek* and has appeared on ZDTV. In the San Francisco Bay Area, she has founded and maintains the Digital Imaging Network (DIN), a support group for people who use digital imaging in their work.

Issues in Digital Archiving

By Bruce Fraser

Building a digital archive is an ambitious undertaking, but one that offers important benefits. It allows institutions to share their collections with a wide audience thanks to the World Wide Web, CD or DVD technology (or their successors), and future digital media we may not yet have envisioned. It offers the possibility of allowing audiences to experience works that are too fragile to put on display, while at the same time creating a permanent record of such works before they deteriorate further. It can serve as the basis for producing facsimiles for eventual sale in a wide range of quality levels. Last but not least, as institutions begin to collect new works that are fundamentally digital in nature, an archiving initiative will be essential if these works are to prove less ephemeral than their analog predecessors.

But if digital archiving is to become a reality, several issues will need to be addressed, ranging from the technical to the philosophical. This piece does not attempt to provide definitive answers. Rather, my hope is to stimulate discussion, provoke critical thinking, and offer some suggested directions for further research and, eventually, implementation.

A digital representation of artwork is a collection of ones and zeroes that, given the appropriate hardware and software, can produce a tangible representation of that piece of artwork. The issues in creating an archival digital representation of artwork revolve around two poles:

- Ensuring the permanence of the ones and zeroes.
- Ensuring that the ones and zeroes produce the best possible representation of the original.

Neither of these is trivial, but the first is less fraught than the second is, so we'll deal with it first.

Data Permanence

Digital data can be thought of as permanent in a Platonic sense: we can copy it without generational loss, each copy being identical to the original. But digital data doesn't exist in a pure Platonic realm: it's stored on some physical medium, and any physical medium is subject to decay and, ultimately, to failure. Worse, while analog media such as photographs decay slowly, failure of digital media is sudden and catastrophic. Technologies such as CD-R and DVD-RAM use built-in error checking that makes them tolerant of media errors up to a point, but once that point is reached, files or whole disks suddenly become unreadable, and in many cases, unrecoverable.

A second related problem is that media go in and out of fashion. Long before the storage medium itself has deteriorated beyond the point of recoverability, the equipment to read it may become scarce, expensive or unavailable. It's still possible to find establishments that can read nine-track magnetic tape, or 8-inch floppy disks, or any of the early, obscure magneto-optical cartridge formats, but the process is inconvenient and expensive. Since the equipment is no longer being manufactured, there will come a time when these storage media become unreadable due to lack of a reader, even if the data itself is still intact.

We also need to address the issue of file formats. Even if the storage medium is readable, and a device exists to read it, the data is useless unless there also exists software that can interpret it correctly, so the choice of file format is also crucial. Ideally, the chosen file format should be widely supported, well-documented in a public specification, and capable of storing uncompressed pixel data with embedded colorimetric information. At present, TIFF (Tagged Image File Format) is probably the most sensible choice, since it's clearly specified and well-supported by a wide range of software packages, but better formats may present themselves in the future.

Therefore, at the risk of stating the obvious, it is imperative that any digital archiving initiative allocates time and resources to allow for periodic refreshing of the storage medium, migrating to new storage technologies and converting to new file formats when appropriate. Otherwise, it's entirely likely that at some point, data will be lost. It's always tempting to divert resources to meet more immediate needs, but in this case, once the need manifests itself, it's too late.

Color Accuracy—The Limits of the Possible

Where the storage issues are relatively clear-cut, ensuring that the data contains the most accurate representation of the original that's possible with current technologies is a much more complex issue that inevitably involves some rather deep discourse as to the nature of color. Some of this discussion is necessarily quite technical, but I'll try to keep the color science arcana to a minimum.

The Visible Spectrum

The major difficulty in trying to discuss color in an objective, quantifiable way is that color is something that really only happens inside our heads. A reasonable working definition of color is that it is

our human response to varying wavelengths of visible light—that part of the electromagnetic spectrum whose wavelengths lie in the range between approximately 380 nanometers and 720 nanometers. (Infrared lies immediately below visible light, and ultraviolet lies immediately above it.)

When we talk about "measuring color," what we measure is the stimulus that our eyes receive, not the infinitely more subtle response we have to that stimulus. A great deal of empirical work has been done on checking the reported human subjective response to different measurable stimuli, and it's clear that there is a strong correlation between the objective stimulus and the reported subjective responses. However, our perception of color is modulated by many factors, ranging from the context in which the stimulus is received, physiological factors such as diet and fatigue, to still more subtle factors such as emotional state—common everyday expressions such as "seeing red" or "green with envy" may contain more literal truth than we sometimes think. In short, quantifying color is necessarily a slippery endeavour, but in this arena, it's also an essential one.

The most unambiguous description of color that we can obtain is the spectral power distribution of a sample, obtained by measuring with a spectrophotometer. These instruments measure the intensity of light present at different wavelengths when light of a known composition is reflected by or transmitted through a sample. Inexpensive spectrophotometers measure light in 10-nanometer or 20-nanometer bands, while more expensive instruments divide the spectrum into narrower bands. Since the spectrum of the light source is known, we can tell exactly how much energy a sample will reflect or transmit at each wavelength. This allows us to predict how the sample will appear under different light sources.

But neither our eyes nor our capture devices work this way. Instead, they use one of a variety of tristimulus color models. A tristimulus color model defines color in terms of the relative amounts of three primaries. This also gives rise to the concept of a color space—when we define a color using three sets of values, we can treat these values as Cartesian coordinates in a three-dimensional space.

Tristimulus Color

Specialized photoreceptors in the human retina form the physiological basis for tristimulus color. We all have two distinct types of photoreceptors, rods and cones. The rods are not sensitive to color, and contribute primarily to our low-light and peripheral vision. Our color vision comes from the cones, of which we have three different types, designated the L, M and S cones (for long, medium and short wavelengths) because each type is sensitive to different wavelengths of visible light. It's this property of our eyes that makes the whole idea of primary colors—colors from which we can make all other colors by combining them in various proportions—possible.

Our digital capture devices—scanners and digital cameras—also use tristimulus color. Typically, they use three sets of photoreceptors filtered with Red, Green and Blue filters, respectively. The Red, Green and Blue filters create a response that's crudely analogous to the L, M and S cones in our eyes, but there are important differences between the two. The L, M and S cones have broad, overlapping spectral sensitivities, while the R, G and B filters used in scanners and cameras are narrow-band filters with spectral responses optimized for the spectral response of photographic dyes. Scanners are generally designed with photographic originals in mind, and while digital cameras are not normally intended to capture photographic prints, they use the existing filter sets that were designed for scanners.

One practical implication of this difference between capture devices and our eyes is that our capture devices (including traditional photographic film) sometimes "see" color differently from the way humans do. Capture devices may distinguish differences between two samples that appear identical to human vision, or they may fail to distinguish a difference between two samples that appear different to us.

Illuminant Dependency and Metamerism

But there's a more fundamental difference between spectral and tristumulus color. With spectral data, we can predict how a sample will appear under any light source whose spectral composition is known. With tristimulus color, we cannot. The corollary of this is that all tristumulus color specifications are specific to a particular illuminant, or light source.

The only way to make two color samples match under all lighting conditions is to ensure that they have identical spectral power distributions. To create a reproduction of a painting that matched the original under all lighting conditions, we would have to essentially repaint the work using the same pigments and canvas used by the original artist. This is obviously impractical.

To create color matches from an original to a reproduction, we rely on a phenomenon known as metamerism, where two stimuli with different spectral responses can appear identical (that is, generate identical tristimulus responses) under a specific light source. Since our eyes and our capture devices use three filters, they don't have a unique response to each different spectrum. Instead, there are many different spectra that can produce the same tristimulus values.

We tend to think of metamerism as a problem—colors that appear to match under one lighting condition no longer do so under another—but we should also bear in mind that metamerism makes it possible to produce, for example, an offset print that recognizably matches an original painting.

At some time in the indeterminate future, we may have capture devices that produce spectral data for each image pixel, which will allow us to create truly unambiguous archival digital color, but that day is far off. (Current spectrophotometers have measurement apertures in the 5mm to 8mm range, which makes them very "low-resolution" devices.) For the foreseeable future, we simply have to do the best we can within the constraints of tristimulus color models.

Device-Specific and Device-Independent Color

Tristimulus color models fall into two distinct classes, device-specific and device-independent. The tristimulus RGB (red, green and blue) captured by today's scanners and digital cameras is "device-specific." What that means is that the RGB values generated in response to a specific color sample will vary from device to device. The basic reason for this is that different capture devices use different filter sets and light sources, so they will produce different RGB values from the same color sample. As a result, RGB color is ambiguous—it has no precise color meaning, just as different people will think of different colors when presented with the word "red." We can (and must), however, take steps to render device-specific

RGB colorimetrically unambiguous by implementing color-management technology. To understand how these technologies work, though, we must first examine the difference between device-specific and device-independent tristimulus color models.

Device-specific color models such as RGB represent color in terms of red, green and blue primaries, but these are rarely the spectrally pure red, green and blue that we can obtain by splitting white light into its components using prisms. RGB color is essentially a set of control signal values that, when sent to a physical device such as a computer monitor, will produce something we then experience as color. But the actual color we perceive will depend on the physical characteristics of the monitor to which the control signals are sent. Monitors produce colors using chemical compounds called phosphors, which emit light at specific wavelengths when excited by a beam of electrons, the amount of light emitted being proportional to the strength of the electron beam. The actual wavelengths emitted vary according to the phosphors used, and these vary from monitor to monitor.

We can think of RGB as being a recipe for color—if we give the same recipe to 20 different cooks, we usually wind up with 20 somewhat different dishes. The same holds true for the CMY (cyan, magenta and yellow) or CMYK (cyan, magenta, yellow and black) models that we use to send to our printing devices. The actual color obtained depends on the color of the individual inks, and the paper or other substrate on which they're printed. So device-dependent color is fundamentally ambiguous with respect to the actual color it represents, unless we take steps to disambiguate it.

To do so, we rely on a different set of tristimulus color models known collectively as device-independent color. Device-independent tristimulus models are not based on physical device control signals. Instead, they're rooted in the spectral response of the color receptors in our eyes, the L, M and S cones. The device-independent tristumulus models in use today were developed by the Commission Internationale de l'Éclairage (CIE), the international standards body charged with developing and maintaining standards for light. Since color is fundamentally a property of light, color falls under the CIE's responsibilities.

There are several different CIE color models, all based on human vision, but CIE XYZ and CIELab are the ones that form the basis of today's color- management technology. By definition, they represent the entire range of perceivable color. They represent perceived color in an unambiguous way, with a caveat. CIE values are only unambiguous when we also specify the spectral power distribution of the light source under which the samples are viewed, because different light sources will produce different tristimulus values for the same color sample. In addition, we need to specify a standard observer. The CIE has defined two standard observers, labeled the 2-degree and the 10-degree standard observers, to address the fact that the cones in our eyes are concentrated in the center of the retina, in an area known as the fovea. We see a color sample that subtends a 2-degree arc centered in our field of vision rather differently than we do one that subtends a wider area and brings more of our peripheral vision into play. So a set of CIE tristimulus values must be qualified by specifying the illuminant, or light source, and the observer. Most uses of CIE colorimetry use a D50 or D65 illuminant with the 2-degree observer.

These perceptually based mathematical models of color vision form the basis of today's color-management technologies.

Color Management 101

Color management systems use software and measurement hardware to provide color consistency across different applications, between our various input, display and output devices. Early color management systems were closed, proprietary systems that locked users into a single vendor, but thanks to the work of an industry group called the International Color Consortium (ICC), we now have an open standard for color management. Almost all the color-management systems available today are ICC-compliant, which means that they are interoperable and platform-independent.

Color management systems, such as Apple Computer's ColorSync and Microsoft's Windows ICM, try to do several useful things. First and foremost, they seek to disambiguate device-specific RGB and CMYK values by relating them to device-independent values, so that we know what actual perceived color a given set of RGB or CMYK numbers is meant to represent. Secondly, they attempt to provide automatic color matching between devices by changing the control values that are sent to each device so that the perceived color remains constant. Third, in those cases where a color present in the original source cannot be reproduced on the destination device due to its physical limitations, they attempt to substitute reproducible colors in some useful way.

Color management systems use three components to perform these tasks:

1. A reference color model, usually CIE Lab or CIE XYZ, which provides an unambiguous description of perceived color.

2. Device profiles, which are basically lookup tables that correlate device-specific RGB or CMYK values with the actual perceived color those values produce on the device in question. A fully-implemented color management system uses profiles for input devices such as scanners or digital cameras, for the monitor (which can act as both an input and output device), and for output devices such as desktop printers, film recorders, and printing presses.

3. A software engine known as a CMM (Color Matching Method) that does the actual work of converting our device-specific values as they travel from device to device. The CMM also handles the task of interpolating the colors that lie between the data points in the device profiles.

When we ask a color management system to perform image matching, we need to supply two profiles, one for the image source and another for the image destination. The color management system then converts the image one pixel at a time. In the classic color management workflow, the CMM looks to the source profile to determine the Lab or XYZ values that correspond to the device-specific color values. Then it converts the device-specific colors to those Lab or XYZ values. As a last step, it looks to the destination profile to determine the device-specific values the destination device needs to reproduce the Lab or XYZ color. In practice, to improve speed and reduce quantization errors, most color management systems use the source and destination profiles to build a lookup table that goes directly from source device-specific color (for example, scanner RGB)

to destination device-specific color (for example, print CMYK), then converts each image pixel through this lookup table.

Color management systems also perform the important task of resolving different devices' color gamuts. Each of our scanners, monitors and printing devices has a range of color it can reproduce. For example, a monitor can't display a red that's more saturated than the red primary. So we need to tell the color management system how we want it to handle out-of-gamut colors—that is, colors present in the source image that are unreproducible by the target device. The ICC specifies four methods of doing so, called rendering intents. These function are as follows:

- Absolute Colorimetric rendering reproduces in-gamut colors exactly, and clips out-of-gamut colors to the nearest printable hue, sacrificing brightness and saturation as necessary.
- Relative Colorimetric rendering is similar to Absolute Colorimetric, with the exception that it scales the white of the source to the white of the destination, and adjusts all other colors accordingly.
- Perceptual rendering attempts to compress the entire color gamut of the source (the range of color represented by the source profile) into the gamut of the destination in such a way that overall color relationships, and hence overall image appearance, are maintained, even though all the colors may shift slightly in the process.
- Saturation rendering maps fully saturated source colors to fully saturated output colors, allowing shifts in hue and brightness. It's primarily useful for business graphics and charts, where vivid color is desired but the actual hues are not important.

Understanding and correctly handling rendering intents is important in all applications, but particularly so in digital archival color. Ideally, we want to capture the entire range of color found in the original work, and in many cases, some of that color will be unreproducible on a particular monitor or printing device. In the graphic arts and printing industry, it's a common practice to limit the range of color captured to the range that we can reproduce, but for archival use that simply isn't appropriate if we want to produce the best digital representation of the original that we possibly can.

While adoption rates vary, color management systems have proven quite successful in the printing and graphic arts industries, including color-critical work such as advertising and fine-art reproduction. They represent our current best hope in creating archival digital color. But they have significant limitations, and allow many different implementations, some of which are more appropriate for archival uses than others. The remainder of this piece focuses on specific issues and tentative recommendations for implementing color management in an archival setting.

Capture

Whether we produce the digital archive by photographing originals on film or by direct digital capture using digital cameras, it's unlikely that our captures will match the original perfectly due to input device metamerism issues. We know that film often captures colors somewhat differently from the way we see them, since the spectral sensitivities of photographic dyes are significantly different from the cone responses in the human eye. The same is unfortunately true of digital cameras, since the narrow-band RGB filters they use are likewise very different from the broadband cone sensitivities. This is not a problem that color management can fix for us. Color management simply describes how our capture devices behave—it doesn't change their inherent shortcomings. So while we can and should build device profiles for our capture devices, these in themselves will not always produce ideal results.

Without a doubt, color management will improve in the future as we gain a deeper understanding of the human visual system. If we want to be able to capitalize on future improvements, we should make sure that we archive the raw captures, along with as much information as possible about the way the image was captured, including, if at all possible, not only the best available device profile for the capture device, but also a record of the spectral power distribution of the light source used for the capture, and a record of the spectral responses of the filter set used by the capture device. Both records should be obtainable from the respective vendors, albeit with some prodding required. The purpose of this archive is to avoid the need to subject originals to the capture process in the future. Some originals may even be too fragile to subject to today's capture processes, all of which involve subjecting the original to a significant amount of light.

The raw capture archive should be kept offline, to await future developments that may improve the final digital representation. Such developments may include better device profiling techniques, or even a better underlying model of human vision. If we know the spectral makeup of the light source used in the capture, and the spectral sensitivities of the filters, we should be able to take advantage of these developments.

The Master Archive

The raw captures serve as the basis for the working archive, the master files from which purposed versions are produced for print or electronic distribution. Some colors in the raw capture will almost certainly be "wrong" in that the digital file may represent accurately the color seen by the capture device, but not by our eyes.

So we need to produce master files by having skilled operators working in a color-calibrated environment edit the raw captures to provide the best possible visual match under the lighting conditions in which the archive will be used. We need to save these files with an unambiguous colorimetric encoding. For simplicity's sake, it's easier to keep our corrected, matched master files in a single standard color space, rather than simply using the native space of each capture device. A second argument in favor of a standard space, and against working exclusively in capture spaces, is that capture devices are often quite non-linear. As a result, the same increment of change in pixel values may produce different visual degrees of change in different parts of the tonal range or color spectrum. For editing, we prefer to use a space that is perceptually uniform, meaning that the same editing increment produces the same amount of visual change anywhere in the color space.

The obvious candidates are CIE Lab, which is designed to be perceptually uniform, or one of a number of abstract, calibrated RGB spaces that have a large enough gamut to encompass the full range of color in the original. Each approach has its pros and cons.

CIE Lab is commonly used as an archiving and interchange color space in Europe, but it has proved much less popular in the

United States. Among its advantages are that it is unambiguous, is supported by the TIFF format, is readily understood by many software applications, and doesn't require a device profile. However, it also has two significant disadvantages. First, it is a very large color space, since by definition it contains all visible colors. This can lead to problems with typical 8-bit per channel encodings, since the 256 possible data values in each channel are spread over the entire visible spectrum, leading to the likelihood of significant quantization errors.

What is quantization error? When we convert from capture RGB to Lab, each RGB value gets mapped to a Lab value. But since both the RGB and the Lab values have only 256 levels per channel, and the Lab values represent a much wider range of color than the RGB ones, we often encounter a situation where two different, but close, sets of RGB values map to the same Lab value. If you like, Lab uses the same number of points to map a much wider territory than the RGB spaces. Hence the distinction between the two original sets of values is sometimes lost in the translation to 8-bit-per-channel Lab.

A 16-bit-per-channel implementation of Lab would solve this problem, but at the time of writing, there is no standard for applications to support 16-bit-per-channel (48-bit) Lab images. This may change in the near future, but a more intractable problem is that Lab is counterintuitive for most operators as an editing environment. Lab encodes colors in terms of three primaries, L* (pronounced el-star), a* and b*. L* represents the lightness of the color sample, a* represents how red or green it is, and b* represents how blue or yellow it is. (Red and green, and blue and yellow, are opponent colors—there's no such thing as a greenish red, or a bluish yellow, because they are mutually exclusive.) While there's a strong cognitive basis for representing colors this way—our perceptual systems seem to translate the cone response into an opponent-color system quite early in the perceptual chain—it isn't by and large a natural way for operators to think about color.

Large-gamut RGB spaces are probably better suited for the purpose. Several software packages support 48-bit encoding for RGB, which increases editability, reduces or eliminates quantization error in color space conversions, and reduces the risk of objectionable posterization or color-banding in output files—sudden jumps instead of smooth gradations.

But most of the commonly used abstract RGB spaces in existence today, such as sRGB, have quite small gamuts that will neither hold the full range of color in originals nor adequately represent the range of color attainable on today's output devices. More specialized large-gamut RGB spaces, such as Wide Gamut RGB (built into Adobe's PhotoShop application as a user option) or Kodak's ProPhoto RGB (loadable into any ICC-compliant application), seem to offer the best solution in the near term.

The process of producing a master file would then be to make the original capture, convert it to our standard RGB color space using profiles for the capture device and the standard RGB space, and edit the file to produce the best possible visual match. It would then be saved as a 48-bit TIFF file, with the profile describing the RGB space embedded in the file. (Abstract RGB space profiles are very small, typically 6K or so, so embedding the profile adds very little to file sizes and makes the RGB values colorimetrically unambiguous.)

This archive would serve as the basis for producing digital files for general distribution, and would also be an important resource to those engaged in serious research. But the master archive files would be too large to be easily sharable with the general public. Instead, distribution files would be made from the master archive, at a resolution appropriate for the use envisaged, converted to the native color space of the output device for which they are destined.

In the case of the Internet, there is no single output device—we have millions of unknown monitors on which the images will be displayed. At this point, probably the best we can do for images intended for the World Wide Web is to convert them to sRGB, which represents an "average" monitor. While sRGB is incapable of representing the full range of color in all originals, using Perceptual rendering in the conversion should produce an image whose overall appearance produces a reasonable visual match to the original within the constraints of the monitors on which they will be viewed.

We see sRGB as a temporary solution, a lowest common denominator: e-commerce is starting to build pressure for a more robust color management solution for the Web as vendors try to sell color-critical merchandise. Once such solutions are in place, we can create new Web images from the master archive that take advantage of the latest developments to produce the most accurate representations possible on the Internet. Self-calibrating monitors are already available at a hefty price premium, but if color-critical e-commerce is to succeed, they must become mainstream, and a mechanism must be put in place to adjust the incoming color to display correctly on the specific monitor on which it's viewed. But such developments will take time, possibly a decade or more.

The master archive would also serve as the basis for print reproductions that may range from postcards to high-quality fine-art books. By preserving in the master archive the full range of color present in the original, we have the best possible basis for creating optimal print reproductions, particularly in high-fidelity print processes that use more than the traditional four CMYK inks to extend the gamut.

Special Needs

Although color management is used in fields as disparate as automobile paints, movies, textiles and plastics, its core constituency is the graphic arts and print publishing markets. One area where this print-centric technology may cause problems is that the standard illuminant used in almost all color-management systems is either D50 or D65, which are the standards for print proofing. Museums and galleries generally, though by no means invariably, use lighting with a somewhat lower color temperature. So it's critical that before we start capturing images, we decide what lighting conditions we want to use as the "reference appearance" of the original. Should it be the way it is displayed in a particular institution? A standard?

A further problem is that today's capture technologies, whether using conventional film or direct digital capture, need to put quite a high intensity of light on the subject, which for some originals would clearly be too destructive. One promising avenue of approach is to take multiple low-light digital exposures and average them together. The CCD (Charge-Coupled Device) sensors used in digital cameras and most scanners suffer a dramatic increase in noise at low light levels. Since the noise is random and the image

is not, averaging tends to reinforce the image, but not the noise. This approach is widely used in astronomical photography, and has also been implemented in some low-end scanners, but as yet, no one has applied it to high-quality scanning-back digital cameras. If the camera vendors are made to see a need, they will likely move to fill it, and low-light digital exposures could prove a very noninvasive technique for dealing with fragile originals.

The challenges are significant. The answers are tentative at best. However, we need to start thinking about the issues and answering the questions as best we can, now. If we can put standards in place that allow for future improvement, we can bring the works to a much wider audience, particularly those that are too fragile to be displayed. These standards will also give us a floor to stand on when we're forced to confront the issues as to just what the work is when pure digital works start to make their way into collections. I hope the foregoing will provoke discussion, raise further questions, and provide, if not goals, then at least directions to explore for the future.

Bruce Fraser
Freelance Consultant and Author
Bruce Fraser (bfraser@amer.net) is a San Francisco-based author, consultant and lecturer in the field of color image reproduction. He is a contributing editor to *Macworld* magazine, *www.macWEEK.com* and *www.creativepro.com*, and writes regularly for many other print and online publications. He is co-author of the award-winning RealWorld PhotoShop series published by Peachpit Press.

Content-Based Image Retrieval:

A New Paradigm for Image Retrieval?

By Colin C. Venters and Dr. Matthew D. Cooper

Introduction

Libraries, museums and archives maintain sizeable collections of original and surrogate images. Both at a national and local level, these repositories are embracing technology to resolve issues of preservation and access by digitizing their image collections [Nail, 1994; Enser, 1995]. Image production and use now routinely occurs across a broad range of disciplines and subject fields. The potential application areas of computer-based image retrieval systems are numerous, ranging from art galleries and museums to remote sensing and earth resource management [Volk Beard, 1991; Gudivada and Raghavan, 1995; Eakins and Graham, 2000]. The current trend of digitizing collections of visually oriented material combined with the prolific growth of computer-generated graphics and image data has resulted in a significant swell in the volume of digital image collections. However, the expertise and techniques for effective image retrieval has not kept pace with the technology of image production [Mostafa and Dillon, 1996]. Recent approaches to the problems of retrieving visual data have focused on the retrieval of images on the basis of their visual properties, so called "content-based image retrieval." This article provides an overview of content-based image retrieval and attempts to answer some questions regarding this new and exciting technology. The questions are: What is content-based image retrieval? How usable is content-based image retrieval technology? Can it provide an alternative solution to the management of image collections? Why should managers of image collections be interested in this technology? Attempting to answer these questions inevitably raises more questions.

Image Retrieval

The successful management of any information resource requires ensuring that the content is exploited effectively and is easily accessible by those users whose information needs it is intended to satisfy [Van Rijsbergen 1979; Wilson, 1981; Salton and McGill, 1983]. However, the issues involved in trying to attain this state with non-alphanumeric data are complex and multifaceted.

Image retrieval has been an active research area for several decades with the most significant contributions from the fields of information retrieval and computer vision [Rui, Y. and Huang, T. S. Chang, S-F]. Initial solutions to manage visually oriented material adopted the traditional information retrieval approach, indexing and classification schemes. This resulted in the development of a number of major subject indexing schemes: Art & Architecture Thesaurus (AAT), ICONCLASS and TELCLASS. The first image retrieval systems followed this traditional information retrieval paradigm approach [Chang and Lee, 1991]. However, these schemes provide a limited solution to the overall problem of image management. For example, the AAT was originally intended to be geographically and historically comprehensive, but an investigation concluded that the scheme was unsuitable for iconographical themes, and was heavily biased toward western art and architecture [Petersen, 87].

The main difficulty of applying indexing and classification schemes arises from an individual's subjective interpretation of an image. It is generally feasible to state what an image consists of in terms of the objects it contains, e.g. dog, woman, tree. With knowledge and subject expertise, the reader can identify objects more precisely, e.g. greyhound, Monica Lewinsky, oak. However, what cannot be clearly stated with any certainty is what the picture means, or even whether it means anything at all [Panofsky, 1968; Shatford, 1986; Svenonius, 1994; Hidderley and Rafferty, 1997]. A more pragmatic problem of applying this approach to an image collection concerns the labor intensive nature of the task. Keefe [Keefe, 89] estimated that to index a small collection of 50,000 slides would take approximately 17 working years. Similarly, The Durham Record Project initially envisioned indexing an average of 200 images a day. In reality, only a maximum of 60 images could be indexed in any one working day [Watson, 1994]. It is widely recognized that this approach imposes constraints on the exploitation of an image collection as an information resource [Enser, 1993]. In recent years, developments have focused on the automatic indexing and retrieval of images by their content.

Content-Based Image Retrieval

Content-based image retrieval provides an alternative approach to that of the traditional information retrieval paradigm. The origins of the method have been attributed to the early experiments conducted by Kato [Kato, 1991] into the automatic retrieval of images by color and shape feature [Eakins and Graham, 1999]. Content-based image retrieval involves a direct matching operation between the feature vectors of the query image and a database of stored images.

The process involves computing a feature vector for the unique visual properties of the image. Similarity is then computed by comparing the feature vector of the query image to the images in the database. The result of this process is a quantified similarity score that measures the "visual distance" between two images, represented by the feature vectors in the feature space. This approach permits the objective analysis of pixel distribution and the automatic extraction of measurements from raw sensory data [Del Bimbo, 1999]. The three

most common image matching feature characteristics are color (global and local), shape and texture. A number of recent publications have surveyed the feature matching techniques utilized in the process of content-based image retrieval, and readers are directed to those publications for detailed descriptions [Del Bimbo, 1999; Eakins and Graham, 1999; Rui, Huang, and Chang, 1999]. The following sections summarize the main retrieval techniques.

Color

Color retrieval techniques allow images to be retrieved on the basis of their global or local distribution of color. Global color features analyze the overall distribution of color within the image for both the dominant color and the variation of color throughout the entire image independent of its location. The position of the color value is not significant. As a result, this feature retrieves a very high proportion of false positives. To compensate for this, local color retrieval was devised. Local color features divide the image into subblocks and then extracts the color features from each block. Image similarity is then based on the similarity of the color distribution. The position of the color value is significant.

Shape

Shape retrieval techniques allow images to be retrieved on the basis of either their boundary (parametric external) or regions (parametric internal). The former uses the outer boundary of the shape, while the latter uses the region enclosed by the object's contours . Shape functions measure the relative orientation, curvature and contrast of lines in an image. The color, position and absolute orientation of the lines have no value. Shape retrieval features are highly sensitive to noise and are most effective when the lines are crisp and clean.

Texture

Texture is a property that displays patterns of uniformity, granularity, directionality and repetitiveness that do not result in the presence of a single color or light intensity [Smith and Chang, 1996]. Several techniques have been developed based on the psychological studies in human visual perception of texture [Tamura, Mori, Yamawaki, 1978]. Texture features analyze areas for periodicity, randomness and roughness of fine-grained textures in images. The color, position and absolute orientation of the features have no value. Texture features can be extremely sensitive to high-frequency features within the image.

Query by Visual Example (QVE)

Content-based image retrieval requires users to submit a query image as the basis of their search. Several user interfaces and their features for content-based image retrieval have been described in the literature, and they have been recently summarized by Del Bimbo [Del Bimbo, 1999]. He reports a number of query tools employed for formulating visual queries by this method, e.g. iconic and by image, painting or sketch. The query by sketch method is used in several systems [Niblack et al., 1993; Del Bimbo, 1997]. Employing this method, users sketch out their desired image by using several features commonly found in computer-based drawing applications. Gupta and Jain [Gupta and Jain, 1997] note that while most systems are limited in the query types they can handle, image query specification should be extended through a range of different tools: query canvas, containment queries, semantic queries, object-related queries, spatio-temporal queries. Whether the current or proposed query tools are effective in aiding the process of visual query formulation and construction is unclear, as there is little empirical evidence to date to support such claims. Research to date is contradictory and the validity of the interaction methods continue to remain untested with real user populations.

Content-Based Image Retrieval Systems

A recent survey of content-based image retrieval systems has revealed more than 70 "systems" in existence [Venters and Cooper, 2000]. The majority of the identified systems were research prototypes developed at university research laboratories. These systems exist primarily to test and demonstrate experimental retrieval algorithms. The systems are generally not available for public use or are shipped in a form unsuitable for use without significant development work. Five commercial products were identified although it should be noted that IMatch is distributed under a shareware license agreement:

- Excalibur Visual RetrievalWare, Excalibur Corp.
- ImageFinder, Attrasoft
- IMatch, MWLabs
- QBIC, IBM Corp.
- VIR Image Engine SDK, Virage Inc.

The technology is shipped in three forms: binaries, Software Development Kits (SDK), and Relational Database Management System (RDBMS) plug-ins.

ImageFinder and IMatch are MS Windows-based solutions and are distributed as binaries. The installation procedure for the software is an extremely simple process, which can be handled by most competent PC users. However, the study noted that there is considerable scope for improvement for both systems in terms of their usability [Venters and Cooper, 2000].

Three SDK's are available: Excalibur Visual RetrievalWare SDK, QBIC Development Kit, and the Virage VIR Image Engine SDK. The SDK's are open application development environments designed to provide system developers with a framework for building content-based image retrieval systems. For example, the QBIC Development Kit includes C++ libraries and header files, four command-line executable modules, source code, a demonstrator application built using the QBIC API, and a mini Web server used by the QBIC demonstrator application, 33 images and reference documentation. There are several O/S distributions for the SDK's including AIX, DEC OSF1, HP-UX, IRIX, Linux, Macintosh OS8 on a PowerPC, Solaris, and WIN95/98/NT/2000.

Both Excalibur, IBM and Virage have integrated their technology into third-party products by collaborating with major RDBMS vendors to extend their products retrieval capabilities to include content-based retrieval. For example, the VIR Image Engine is deployed as a data cartridge for Oracle 8i Server (Enterprise Edition), a datablade module for Informix Internet Foundation 2000 server, as well as being integrated into databases from Sybase, Object Design and Objectivity.

The products support a wide range of image graphic file formats including several industry-standard raster and vector formats: BMP (OS/2, Windows), DDIF, EMF, GIF, JPEG, MAC, PCD, PCT, PCX, PGM, PICT, PNG, PPM, PSD, RLE, SGI, SCT, TGA, TIFF, WMF. The price of the commercial products ranges from $50 (IMatch) to $50,000 (Excalibur Visual RetrievalWare SDK).

The survey established that there are similarities between the commercial and prototype systems in terms of the features they support, i.e. color, shape and texture retrieval. What was not apparent is whether there are considerable differences between the effectiveness of the matching features, and if so, what those differences were. It is also unclear how widely this technology is being adopted and utilized in either the private or public sectors. The QBIC engine is currently being tested at a number of sites including the Russian State Hermitage Museum, the University of California Davis, the Fine Art Museum of San Francisco, and the French Ministry of Culture. The preliminary results of these tests are inconclusive [Holt and Hartwick, 1994]. A study is currently underway at Manchester Visualization Centre to test the feasibility of using current content-based image retrieval systems for the UK Higher Education community. The final results of this study will be published by the fall of 2001.

Content-Based Image Retrieval and User Studies

The major focus of the field has been the design, development and evaluation of robust retrieval algorithms. In comparison with the field of text retrieval, very little is known about the information needs and behavior of end users. Eakins and Graham [Eakins and Graham, 1999] identify and summarize several user studies which focus on specific image collections and user populations. They conclude that while the results provide a valuable insight into the expressed needs of the specific end-user population, they are too limited in their scope to extrapolate the findings to a general theoretical framework. While accepting the limitations of these studies, Besser actively encourages this approach as he suggests that segmented studies will eventually lead to a richer and more detailed understanding of image use [Besser, 1995]. A study is currently underway at the Institute for Image Data Research, VISOR (Visual Information Seeking Oriented Research), which aims to investigate the information-seeking behavior of users with respect to image data. The overall aim of the study is to provide a holistic framework of users information-seeking behavior. Preliminary findings suggest that there is little significant difference between the information-seeking behavior of users with text-based information needs and users with image-based information needs.

Summary and Conclusion

The dramatic rise in the volume of digital images has exacerbated the complex issues surrounding the effective retrieval of both dynamic and static images. This has proven to be a catalyst for resurgence in the field of visual information retrieval. Recent developments in the field have focused on the retrieval of images by their content. Despite the many criticisms leveled at existing indexing and classification schemes, they provide a stable and solid foundation for retrieving relevant items. The framework they provide for image management should not be underestimated. Content-based image retrieval is a fledgling technology, which provides an alternative approach to metadata solutions. It can be viewed as a complementary extension to the traditional information retrieval paradigm. The capability of the technology has been recognized for many years, but has not been fully exploited yet. There is a growing realization in the research community that the current approach to the challenges of retrieving similar visual data needs radical rethinking.

Content-based image retrieval potentially creates new and exciting opportunities for users to query visual data. For example, a user may be interested in how Picasso's usage of color influenced contemporary painters. It is still unclear, however, what exactly these user's questions are. It may be worthwhile to exploit the current state of content-based image retrieval technology by combing it with existing approaches and creating a hybrid system. Only by creating the environment for the user to explore will its full potential be finally realized. Content-based image retrieval is an untapped resource, which is currently limited more by our imagination than by our technology. Further information on content-based image retrieval will be available at *www.cbir.org.uk* in the near future.

Acknowledgements

The authors would like to thank the Joint Information Systems Committee (JISC) for funding content-based image retrieval research at Man- chester Visualization Centre under a JISC Technology Applications Program (JTAP) award.

References

Besser, H., Getting the Picture on Image Databases, *Database*, April/May, 1995, pp. 13-19

Cawkell, A. E., Picture Queries and Picture Databases. *Journal of Information Science*, 19, 1993, pp. 409-423.

Chang, C. C., and Lee, S. Y., Retrieval of Similar Pictures on Pictorial Databases. *Pattern Recognition*, 24, (7), 1991, pp. 675-680.

Cochrane, C., An Overview of Trends in the Collection and Use of Moving Images an the United Kingdom. *Journal of Documentation*, 49, (3), 1993, pp. 278-291.

Del Bimbo, A., *Visual Information Retrieval*. Morgan Kaufmann. 1999.

Eakins, J. P., and Graham, M. E., *Content-Based Image Retrieval*, JTAP. 1999, pp. 1-63.

Enser, P. G. B., Query Analysis in a Visual Information Retrieval Context. *Journal of Document and Text Management*, 1, (1), 1993, pp. 25-52.

Fidel, R., *Challenges in Indexing Electronic Text and Images*. ASIS Monograph Series. American Society for Information Science. 1994, pp. 3-306.

Gudivada, V. N., and Raghaven, V. V., Content-Based Image Retrieval Systems, IEEE Computer, September, 1995, pp. 18-22.

Gupta, A. and Jain, R.., Visual Information Retrieval, *Communications of the ACM*, 40 (5), 1997, pp. 71-79.

Hidderly, R. and Rafferty, P. Democratic Indexing: An Approach to the Retrieval of Film. In: *Proceedings of Library and Information Studies: Research and Professional Practice. 1997. Queen Margaret College*, Edinburgh,Taylor Graham. 1997.

Holt, B. J., and Hartwick, L. Visual Image Retrieval for Applications in Art and Art History. In: *Proceedings of Storage and Retrieval for Image and Video Databases II. 1994. San Jose, California, USA*, SPIE, 1994.

Kato, T. Database Architecture for Content-Based Image Retrieval. In: *Proceedings of Society of the Photo-Optical Instrumentation Engineers: Image Storage and Retrieval, 1662. 1992. San Jose, California, USA*, SPIE, 1992.

Mostafa, J. and Dillon, A. Design and Evaluation of a User Interface Supporting Multiple Image Query Models. In: *Proceedings of the 59th Annual Conference of the American Society for Information Science: Global Complexity, Information, Chaos and Control. Baltimore, Maryland, USA*, American Society for Information Science. 1996.

Niblack, W., Barber, R., Equitz, W. Flickner, M. Glasman, E. Petkovic, D. Yanker, P. Faloutsos, C. Taubin, G. The QBIC Project: Querying Images By Content Using Color, Texture and Shape. In: *Proceedings of Storage and Retrieval for Image and Video Databases. San Jose, California, USA*, SPIE, 1993.

Panofsky, E., Perspectives as Symbolic Form, Zone Books, 1991.

Rui, Y. and Huang, T. S. Chang, S-F., Image Retrieval: Current Techniques, Promising Directions, and Open Issues. *Journal of Visual Communication and Image Representation*, 10, (1), 1999, pp. 39-62.

Salton, G., and McGill, M. J., *Introduction to Modern Information Retrieval*, Computer Science Series, McGraw-Hill Book Company, 1983

Shatford, S., Analyzing the Subject of a Picture: A Theoretical Approach. *Cataloging & Classification*, 6, (3), 1986, pp. 39-62.

Smith, J. R., and Chang, S-F., Automated Binary Texture Feature Sets for Image Retrieval, In: Proceedings of ICASSP'96, Atlanta, GA, USA, 1996.

Svenonius, E., Access to Nonbook Materials: The Limits of Subject Indexing for Visual and Aural Languages. *Journal of the American Society for Information Science*, 45, (8), 1994, pp. 600-606.

Tamura, H., Mori, S., Yamawaki, T., Texture Features Corresponding to Visual Perception, *IEEE Transactions on Systems Man, and Cybernetics*, SMC-8 (6), 1978.

Van Rijsbergen, C. J., *Information Retrieval*. 2nd edition. Butterworths, 1979.

Venters, C. C., and Cooper, M. D., *A Review of Content-Based Image Retrieval Systems*, JTAP, 2000

Volk Beard, D. V., Computer Human Interaction for Image Information Systems. *Journal of the American Society for Information Science*, 42, (8), 1991, pp. 600-608.

Watson, M. T*he Durham Record Project*, BLR&D, 1994

Wilson, T. D., On User Studies and Information Needs, *Journal of Documentation*, 37, (1), 1981, pp. 3-15.

Colin C. Venters
Research Associate
Manchester Visualization Centre
University of Manchester
Colin C. Venters (*c.venters@man.ac.uk*) is a research associate at Manchester Visualization Centre, Manchester Computing, University of Manchester. He is currently working on a two-year project, which aims to assess the feasibility of content-based image retrieval technology for the UK Higher Education community.

Dr. Matthew D. Cooper
Research Project Manager
Manchester Visualization Centre
University of Manchester
Dr. Matthew D. Cooper (*matt.cooper@man.ac.uk*) is a research project manager at Manchester Visualization Centre, Manchester Computing, University of Manchester. Current projects include research and development of image-processing methods, parallel computing applications, and methods and applications for data visualization in a range of disciplines.

MISSION STATEMENT

The Museum Computer Network (MCN) is a nonprofit organization of professionals dedicated to fostering the cultural aims of museums through the use of computer technologies. We serve individuals and institutions wishing to improve their means of developing, managing, and conveying museum information through the use of automation. We support cooperative efforts that enable museums to be more effective at creating and disseminating cultural and scientific knowledge as represented by their collections and related documentation.

Do It Once...
...Do It Right!

Wm. Hogarth, Rake's Progress (detail)
Courtesy of University of California,
Berkeley Art Museum

Digital imaging projects are a major investment in time and resources. It makes good sense to create a high-quality "master" or archival digital image when the time comes to digitize a collection.

You will only handle the art once, yet have a color-accurate, information-rich master image that can be used to generate derivative files. Create duplicate images for web access, database thumbnails, poster printing or for detailed research...without rescanning or photographing the original again.

The Better Light scanning camera system can generate RGB TIFF files up to 549MB for flat art copy work or photography of three-dimensional objects. It is one of the most affordable and versatile digitization solutions for museums, libraries, and historical societies.

High Performance
Digital Scanning Inserts
for Superior Images in
Commercial, Catalog and
Reprographic Applications

- **Exceptional image quality**
- **Unequaled color accuracy**
- **Use with any 4"x5" camera**
- **Fastest "real time" scans**
- **File sizes up to 549 MB**
- **Four models from $10,000**

BETTERLIGHT
...simply better digital images

Visit our website www.betterlight.com

1200 Industrial Rd. #17, San Carlos, CA 94070 • Phone: 650-631-3680 • Fax: 650-631-2915 • Email: sales@betterlight.com

www.willo.com

(e there or e^2)™