25 Years of Digital Archives of Visual Materials: What we've done, What we've learned, & What challenges remain

Howard Besser
New York University Professor of Cinema Studies
Director, Moving Image Archiving & Preservation Program

• Visual Material
  – capturing what the eye sees (or can't see)
  – theories of looking at/seeing images
• Major milestones in Digital Visual Archives
• What problems we've solved; what remain as challenges
  – Issues of Copying and Re-contextualization
• The Larger Questions

Images in Science
— what unaided eye can't see
Both Professionalism & Commercialism

Photography as industry/commerce

Image meaning changes w/context
Major milestones in Digital Visual Archives-
• 1986 Berkeley ImageQuery
• 1992 Montreal Gates Project
• 1995 Museum Educational Site Licensing Project (MESL)
Thumbnail Browse
Anthropology

Boolean Queries

Boolean Queries w/pulldown

 Thumbnails linked to record-list, enlarged image, full record

“Add to Save List” (basket)

Separate integrated image viewing tool
Enlarging Manuscripts

Image Processing to improve legibility

Color Analysis & Manipulation

Color Analysis & Manipulation

Geo-Referencing

Thumbnails linked to icons on map
Any basemap can have full cataloging.

Image Processing (pseudo-color) can be used for research.

ImageQuery features

- Network-accessible
- Client-server
- GUI
- Boolean searches
- Thumbnails linked to text and image
- Separate DBMS
- Tools for annotation, conservation, research
- Browse and sorting
- Geographic elements

1986 ImageQuery Technical Challenges (since resolved)

- Scanning difficult and expensive
- Storage and network constraints (analog videodisc advocates)
- Lack of consistent client software
- Color display maximum of 8-bit
- Lack of compression standards
- Lack of metadata standards

How well did we really understand the problems in 1986? (1/2)

- Boolean Queries
- + Thumbnails linked to text & larger images
- ∣ Client-Server
- - Unix clients
- + Separate the DB from the User Interface
- ∣ External operator tools
  - ∣ right
  - ∣ partly right (we didn’t really understand the problem)
  - - wrong
  - 0 didn’t think of

How well did we really understand the problems in 1986? (2/2)

- ∣ Integration with other information
- ∣ Interoperability
- ∣ Universal user interface to access different types of collections in a similar way
- - Scaling Up
- ∣ Longevity
- ∣ Most users are not from the core domain
  - ∣ right
  - ∣ partly right (we didn’t really understand the problem)
  - - wrong
  - 0 didn’t think of
1992 Montréal “Gates” Project
—3D fly-over (1:52 research-cannons, 2:30 fly-over)

3D ground navigation

Plan view w/DB queries

Slices of time

1992 Gates Challenges
(now easier)

• Difficulties of joining together different text DBs—birth/death records, property ownership, plans & maps
• Lack of mapping tools, geo-location standards
• Processing power
• Lack of visualization tools
• Poor touch-screen interfaces
• No serious preservation plan

1995 Museum Educational Site Licensing Project (MESL)

• Aggregation of 10,000 images and rich metadata from 6 Museums and Library of Congress
• Provide access on 7 university campuses, each using own interface and infrastructure
• Sponsored by Getty AHIP
Fields in MESL Data Dictionary (1.1)

- 1. data agreement number
- 2. holding institution
- 3. accession number
- 4. accession method
- 5. credit line
- 6. label
- 7. object type/objectclass/object name
- 8. object title/caption
- 9. creator/maker - name
- 10. creator/maker - culture/nationality
- 11. creator/maker - role
- 12. creation place
- 13. creation begin date
- 14. creation end date
- 15. creation technique/method/process
- 16. material/medium
- 17. support
- 18. dimension/extent/unit
- 19. parts/pieces
- 20. marks/inscriptions
- 21. object type/objectclass/object name
- 22. dimensions
- 23. concept/topic
- 24. concept/topic-period
- 25. concept/function
- 26. description
- 27. accompanying image - file name
- 28. accompanying image - caption
- 29. accompanying image - capture data
- 30. accompanying document - file name
- 31. accompanying document - type
- 32. version identification
-101. data agreement number
-102. creator
**MESL Challenges**

- Metadata diversity (now easier, but still not solved)
- Image quality standards for aggregation (now easier, but still not solved)
- Tools to make it easier for teachers and students to use these (mostly solved for discovery; not solved for integration into coursework)
- Test use/abuse and develop model "terms of use" (somewhat solved except for © issues)
- Make all stakeholders more comfortable with the circulation/use of these digital images (solved except for some rightsholders)

**MESL-related Issues raised but not solved**

- Scalability for metadata creation for large photographic archives w/collection-level (not item-level) metadata
  - at 1995 RLG Digital Image Access Project meeting, my idea of user-contributed metadata was marginalized
- Economic sustainability of technological infrastructure needed for digital visual DBs

**Major developments post-1995**

- Widespread adoption of Web
- Acceptance of advertising-based business models for organizations outside the public sector
- Dublin Core
- Technical Imaging Metadata Standard
- Ubiquity of digital cameras and scanners-
  - rise of general interest in DIY, and of aggregators wanting to distribute DIY content
Major developments post-1995

- Widespread adoption of Web
- Acceptance of advertising-based business models for organizations outside the public sector
- Dublin Core
- Technical Imaging Metadata Standard
- Ubiquity of digital cameras and scanners-
- Rise of general interest in DIY, and of aggregators wanting to distribute DIY content-
What issues still remain?

• Short-term (likely to be solved reasonably well in near future)
  – Metadata standards and mapping between ontologies
  – Still image capture and compression standards and tools
• Mid-term
  – Economic sustainability for handling large-scale visual archives
  – Acceptance of user-generated metadata grouped by contributor
  – Digital Preservation
• Long-term/Major-
  – Versions and variant forms, Integrity
  – © and Terms of Use that recognize need for Recontextualization

Variant Forms/Integrity

• Which version is authentic?
• How do we (even verbally) distinguish between the same root image stored
  – in different file formats? or
  – at different resolutions? or
  – compressed differently?
• How do we tell whether or not 2 images are from the same family?
Variability of Digital Works

- Works that use external websites as they change over time (I want you to want me)
- Works that compose “on the fly”
- Interactive Works
- Works that rely upon software that will eventually become obsolete

© Issues

- Laws are made for high-value commercial works, yet rules apply to all works
- Little (though recently increasing) recognition of exceptions for scholarly and research purposes
- Underlying Rights
- Little legal recognition of need for Recontextualization

Underlying Rights

- Archives may have a right to the whole but not have a right to the components (such as a copyrighted work in the background)
- Third party information, underlying rights are often granted for a limited time or a particular kind of use, or a particular geographic region. This is particularly salient for multimedia works.
- Eyes on the Prize, I-early 1987; II 1990
  - became unavailable due to lapsed underlying rights by mid 90s
  - Jan-Feb 2005 Eyes on the Screen civil disobedience
  - Aug 2005 Grants for clearing rights ($600K Ford, $250K anonymous) announced
  - Needed to clear video from 82 archives, 273 stills from 93 archives, 120 song titles, many of which weren’t able to be cleared (so were either silent, or had substitutions)
  - $850K was just for fees for content-holders, and didn’t cover all the research, writing, begging, legal wrangling…

Archive could be sued for displaying Shepard Fairey poster

- This is not only a problem for film/video producers, but also for archives & exhibitors
  - If the producer relies upon “fair use” or “public domain” instead of clearing ©, everyone downstream insists upon indemnification
  - A distributor won’t handle a work that hasn’t explicitly cleared every tiny © hurdle
  - An exhibition venue (a broadcaster or even an archive) will be reluctant to exhibit a work unless every single sub-piece has cleared ©
  - A funding agency often won’t give resources to an archive wanting to restore a work unless they’re convinced that it can be shown or copied without © restrictions
  - Even professional organizations insist that we sign a contract stating that we’ve cleared © for every little thing we want to show as part of a conference Talk
My class assignments using Recontextualization-

http://besser.tsoa.nyu.edu/howard/repurposing.html
That Girl From......

I can't believe I'm back here again. The word haunt us about as much as the mystery of her face......

(...) (Continued)

Hafiz had been working on a project that was going to be completed soon. He was excited about it because it would allow him to move from his current job to a more challenging one. However, due to unforeseen circumstances, the project was cancelled. Hafiz felt upset and decided to take some time off to think about his next steps. He went for a walk in the park and found a bench to sit on. It was a beautiful day, and the park was filled with people enjoying the weather.

The T-Shirt Bandit

In the popular UCLA professor internes T-shirt debate

UCLA arts and (Continued)
20th century creativity built upon prior Works
new works involve repurposing/recontextualizing old

- Collage
- jazz
- Pop
- Postmodern
- Rap/Sampling
- Re-Mix, Mash-Ups

Picasso’s Violin (1912)

In last 50 years, people certain elements in society have relabeled “homage” as “Stealing”

Kruger’s Untitled 1990 (It’s a small world but not if you have to clean it)

Fox takes Franken to court; they think they own “Fair & Balanced”

Fox takes Franken to court; they think they own “Fair & Balanced”

From University of Florida clarinet studio web site
Http://www.arts.ufl.edu/music/clarinet/images/mona-lisa.gif

From a web page devoted to Star Wars humor
Http://humour.naboonline.com/images/monalisa.jpg

From web site for a head shop in the United Kingdom
Http://www.the-head-shop.co.uk/images/The-Original-Mona-Lisa.jpg

Mona Lisa Socks, For sale online at http://www.studiolo.org/Mona/MONASV13.HTM#LIST

Mona Lisa Bathing Suit, For sale online at http://www.studiolo.org/Mona/MONASV13.HTM#LIST
Fox News calls documentary *OutFOXed* “illegal copyright infringement”

Viacom sues YouTube for $1 Billion

ABC makes Spocko’s ISP shut down his blog because he posted racist ABC video clips

“This Film IS Not Yet Rated” feared being censored for using MPAA clips

EMI and Sony try to shut down DangerMouse’s re-mix of vocals from Jay-Z’s *Black Album* and Beatles’ *White Album*

Copyright Censorship

- Material is sometimes withheld from producers for reasons of politics or embarrassment
  - NYT refusing © permission for Michael Moore reprint their Iraq coverage editorial apology
- Some material never makes it into productions because © is too expensive or difficult
- The current copyright regime imposes censorship in the form of “prior restraint”
Like Scientific Info, Art & Popular Culture Images are part of our heritage
- Raw material for public discourse
- Appropriation is important for a democracy and a Civil Society

Today, images are infinitely reproducible
- Rightsholders try to use the Law to protect their work's
  - Perceived Aura
  - Authenticity
  - Uniqueness
- And to make reproduction or recontextualization illegal so that they can maintain their monopoly

Legal Actions against recontextualization

Legal actions against Satire

Nervy to sue others for appropriating its intellectual property
Historically, copying of Art was viewed as *homage*, or at least as perfectly acceptable commentary.

Madonna & Child

Near future: How are things changing as:

- Everyone has a camera everywhere they go
- Huge numbers of people have access to easy-to-use tools for manipulating images
- Distribution of images worldwide is easy, quick, and cheap

21st Century creativity, teaching, and commentary will depend even more on recontextualization (Re-Mix)

- George Bush Don’t like black people (Black Lantern)

Large Issues we’re now confident about

- For the non-specialist, Discovery will be through aggregation services. Our archives will continue to service Discovery either by ourselves or collectively.
- For the non-specialist, the importance of context is disappearing
Larger Questions that have been with us since the beginning

- Will people continue to visit archives & museums?
- Will surrogates substitute for original objects?
- Tension btwn Democratization & Aesthetic contextualization (connoisseurship)
- Does Democratization imply commercialization?
- What will be the right balance btwn structured data and key-word tagging?
- For which corpuses of images will content-based indexing and retrieval be most useful?

25 Years of Digital Archives of Visual Materials:
What we’ve done, What we’ve learned, & What challenges remain

- http://besser.tsoa.nyu.edu/howard/Talks
- http://www.nyu.edu/tisch/preservation/
- http://besser.tsoa.nyu.edu/howard/repurposing.html