Context Recombination for Cultural Digital Archives

KITAMOTO Asanobu
National Institute of Informatics

http://dsr.nii.ac.jp/
National Palace Museum, Taipei
National Palace Museum, Taipei
Participating in Time

A Western painting is usually finished after the artist has stopped working on it. However, in the artistic and cultural environment of China, the surface of a painting after completion can accumulate many additions by artists, viewers, and collectors over the years. The result is a "collective" work that transcends both time and place. You will often find seals and inscriptions on many of the paintings on display here. These markings record a wealth of information about the works, such as who appreciated, viewed, and collected them. It is also important to note that in the Chinese tradition, marking a masterpiece is not considered a form of defacement, but a method of veneration and pedigree. Passing through time and different hands, a work of art takes on a life of its own, continually aggregating visual and textual records that change its original appearance. Consequently, in addition to appreciating the original painting, the placement, inherent beauty, and record of seals and inscriptions are also important focuses. Inscriptions and seals not only take us into a dialogue between the ancients and paintings, they also record many specific events that occurred over time and space.

National Palace Museum, Taipei
Collective Work

- After completion, **the surface of a painting** can accumulate **many additions by artists, viewers, and collectors over the years.**

- **Markings** record **a wealth of information about the works**, such as who appreciated, viewed, and collected them.

- The result is a **collective work** that **transcends both time and space.**
Digital Archive is also a Collective Work

• After digitization, each data can accumulate many additions by artists, viewers, and collectors over the years.

• Metadata record a wealth of information about the works, such as who appreciated, viewed, and collected them.

• Digital archive is a collective work that transcends both time and space.
Collective Annotation

This statue is from Turfan.

Statues made in 8th to 9th Century in Central Asia

An older work

This statue is from Kuqa.

Wheels are important to represent the identity.

The style of hands is very interesting

• Annotation graph is a logical structure that creates contexts for interpretation.
Ambiguity Problem

- A single image allows many interpretations.
- How a viewer can be inspired to discover new interpretations?

Context plays an important role!
Long Term Goals

• The digital archive is not only the repository of digital resources but also a place for communication and interaction over them.

• It leads to the enhancement of our knowledge through the discovery of new interpretations, understandings with the help of context.
The Model of Content

- Content = Context-free content + Context-sensitive content
- Context-sensitive content = Relationship + Arrangement

- Context-free content is the basis of CBIR.
- Context-sensitive content is important in relation to ambiguity problem.
Content-Based Image Retrieval

Image features to find images that have similar contents with the query.

• Can we define the content of images with their own right?
• What is the real use of similarity?

What we really need is quick browsing and comparison over many images. CBIR is a tool for this task.
Context-based Image Navigation

• Searching (retrieval) is only a part of the whole story.
• Image navigation is repeated interactively to reach interesting contexts for a viewer.
• A viewer may recombine some resources to create a new context.
Context Recombination

• “Context” represents any kinds of relationship among a set of data.
• The interpretation of data depends on in which contexts the data is presented to a person.

A database system that supports the recombination of context at user’s will.
A Context by Arrangement

• The arrangement of images gives rise to a context from the proximity of images.
• The rearrangement of images leads to another proximity and ordering.
Another Context

- The visualization of logical relationship between resources in a high-dim space.
Yet Another Context

Buddhism Art

Statue

Devatä-Halbfigur

Wall Painting (Fragment)

Zenitstreifen des Deckengewölbes

Baumgottheit

• Taxonomy and thesaurus creates another context useful for understanding.
Rethinking about Databases

**Traditional Thinking...**

- Databases are for searching information.
- The result is a set – the arrangement of data usually doesn’t matter.

**New Way of Thinking...**

- Databases are for creating contexts.
- The result is an arrangement – a new context leads to a new interpretation.
Query Language

• This amounts to the development of a query language at the system level.

• SQL is THE giant in relational databases and is still expanding.

• SQL is not a perfect language and focuses on searching, hence little support for creating “contexts.”
Digital Silk Roads Project (1)

• Started in 2001 supported mainly by the government funding agency.
• Agreement with UNESCO toward scientific collaboration.
• Two symposiums and research collaboration with Asian researchers.
• Project web site - http://dsr.nii.ac.jp/
Digital Silk Roads Project (2)

Several Sub Programs

- Bamiyan Virtual Museum
- Caravanserais along Silk Roads
- Dunhuang Field Research

Focused Research Areas

- Image Processing / Analysis
- Geographic Information Systems
- Content Management / Databases
- Metadata / Ontology
NII Digital Archive Projects

- A set of relatively independent projects.
- Digital archive projects receive high priority in NII.
Toyo Bunko (Oriental Library) Digital Archive

• “Morrison Library” : Books about Asia written in European languages.

• Comprehensive collection of basic books for Asian study.

• Accessibility unsatisfactory because of the preservation of fragile books.

• Our project: digitize those books 2,000-4,000 pages each year for public access.
Accessibility

• Enhance accessibility to old books for everyone in the world through...
  1. Digitization
  2. Multimedia processing
  3. Multilingual text processing
  4. Database technology
  5. Internet and Web
  6. Collaborative work
Search into the Book

1. Optical character recognition (OCR) is applied to every page of the book.

2. Machine translation is applied to OCR text.

3. Manual text input is also used.

4. A full text search engine is used to search into the book.
Web Site

• Over 5600 pages from 27 rare books.
• Resolution = 4000x4000 or 10000x12000.
• Built on **open source software** (Linux, apache, perl, postgreSQL, libxml2).
• **XML and XSLT** is intensively used.
• The internal server and the public server is synchronized by rsync.
• System developed by myself 😊
• **Publicly available by the end of March!?**
Summary

• The ambiguity issue is important in the cultural domain.
• We proposed the model of content as the combination of context-free and context-sensitive elements.
• “Context recombination” was proposed as a unifying principle.
• Demonstrated our digital archive.