

Past Forward - Digital Media for Cultural Heritage: The Case of the Global Memory Net¹

Ching-chih Chen

Professor

Graduate School of Library and Information Science

Simmons College, Boston, MA 02115, USA

chen@simmons.edu

Abstract:

Global Memory Net, a major International Digital Library Project of the US National Science Foundation has enormous potential for resource sharing as well as collaborative research and development. This paper extends the keynote speech delivered in Delhi on digital content and community building for archiving and sharing cultural and heritage resources globally at the ICDL meeting in Delhi on February 26, 2004. It will focus on the recent collaborative efforts in research and development. Specifically, discussions will center on multimedia digital library development, image retrieval from traditional metadata based to content-based image retrieval (CBIR), the plan for more audio and digital video handling, the use of Informedia technology for digital videos, and the collaborative Project Restore, in which images on artifacts restored by using nanoparticle technology are included as a part of GMNet. GMNet also offers a perfect testbed for further collaborative cutting-edge technology development.

1. INTRODUCTION

At the *International Conference on Digital Libraries (ICDL)* meeting in Delhi three days ago, I was honored to deliver an invited keynote speech, entitled “GLOBAL MEMORY NET: Potential and Challenges for Archiving and Sharing Cultural and Heritage Resources” [Chen, 2004]. In that paper, I discussed both the process related to the need for building of a global digital library via community and infrastructure building across national borders.

Today I am delighted to deliver the 10th *Annual Meeting in Informatics* in Bangalore, and have decided to extend my presentation to place much more on the need of technological issues and the need for collaborative efforts in that area. It is important to keep in mind that *GMNet* is heavily involved with multimedia contents, thus most of the collaborative R&D activities have centered in the use of these contents.

2. FROM CMNet TO GMNet

CMNet is an International Digital Library Project supported by the US National Science Foundation since 2000. It is intended to develop a model for international collaboration with various R&D activities in digital libraries. It hopes to accomplish “more” with “less,” avoid duplication efforts, and capitalize R&D results from other funded digital library projects. In the short three years, it has made a great effort in developing collaborative infrastructure for digital library development and actual multimedia digital content development.

¹ This is an invited talk given at the 10th *Annual Lecture in Informatics* in Bangalore. The paper will be given in a very interactive way with considerable multimedia presentation as well as real-time online retrieval. It is impossible to include all those in this paper. For this reason, I hope that the readers of this paper will use this paper as a summary of what is actually presented.

CMNet's (Figure 1) core contents build upon the large quantity of visual materials of the author's earlier interactive videodisc project supported by the US National Endowment for the Humanities, entitled *PROJECT EMPEROR-I* with award winning interactive videodisc (1989) as well as multimedia CD (1991) products, both called *The First Emperor of China*.



Figure 1. CMNet Home Page

Figure 2 shows that it is related to one of the most significant historical period of China. The site is one of the most significant archaeological find of the 20th century. To each image included, extensive research efforts



Fig. 2. First Emperor of China's warriors

were made to provide relevant descriptive data (metadata) with annotations, as well as links to relevant references and texts whenever possible. This labor-intensive research has paid off because these data has become the attractive basis for a number of exciting and productive technology-oriented collaborative works, to be discussed in the following.

Because of the initial success of the project, *CMNet* has grown naturally to *Global Memory Net (GMNet)* in order to involve more valuable contents in the world as well as more subject areas, since the technological tools developed to organize, retrieve, and present sought-after images and videos can be used for all subject contents. Figure 3 shows only a small sample of additional types of contents:



Figure 3. Some sample cultural and heritage contents in *GMNet*

3. COLLABORATIVE R&D ON DIGITAL MEDIA FOR CULTURAL HERITAGE

At the time when we proposed the *CMNet* project in 1999, digital library initiative has already started for a number of years in the US. While considerable progress has been made in numerous high-power focused R&D research, there were few integrated functional libraries with large volume of digital contents in multimedia formats,” and there was not enough digital libraries connected together from distributed environments. This is why *CMNet* and now *GMNet* have placed emphases on multimedia content development and management [Chen 2003]. Putting that aside, I shall share with you the other aspect of activities which stress the need for collaborative R&D in addressing other important problems, specifically our collaborative work with the cutting-edge technologists in capitalizing the new tools and techniques already developed in their funded projects and in furthering their research by using the rich multimedia contents and associated metadata. In this regard, we will show the importance of cross-disciplinary collaboration in large-scale global digital library development [Chen 2001, Chen 2001a, Figure 4].



Fig. 4. Chen's 2001 book

3.1. OAI Research Among Chinese Collaborators

Each of our collaborators in China has dealt with different “memory” related contents:

- ✚ Tsinghua University (Beijing, China) in Chinese architecture,
- ✚ Peking University (Beijing) in rubbings, and
- ✚ Shanghai Jiao-Tong University in Chinese music and musical instruments.

The difference in subject matters has resulted in using different metadata schemes although they all seem to be in compliance with Dublin Core. Therefore among them, under the collaborative funding with *CMNet*, they have developed a common OAI method to harvest the metadata of each institution for interoperable use of these cultural resources, and are currently testing the use of the extensive metadata of the Emperor images. The metadata scheme of Emperor is deliberately different from the standard one for the purpose of doing more cross-mapping research.

3.2. Intelligent Agent and Image Retrieval

Since 2000, the collaborative research with Prof. V. W. Soo of the Computer Science Department of the National Tsinghua University has yielded interesting results in developing similarity matching algorithm for retrieving relevant Emperor images (Figure 5). The preliminary Emperor's metadata have been used to develop ontological terms (Soo 2002; Soo 2003), and more extensive data will be used. Because the ontological terms are mixed with Chinese characters, I shall refer those of you who are interested in this subject to the published papers.

Matching a query schema with annotated images.

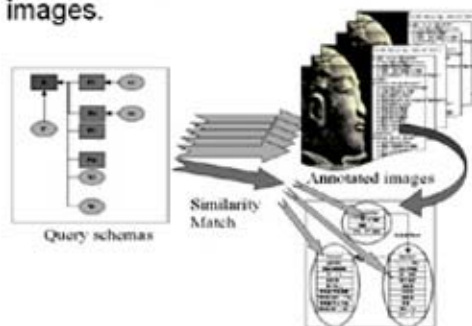


Fig. 5. Soo's work using Emperor metadata

3.3. Semantic Sensitive Content-based Image Retrieval

Since 2000, close collaboration with Prof. James Z. Wang of the School of Information Sciences and Technology of the Pennsylvania State University has enabled us to explore the use of semantic sensitive content-based image retrieval (CBIR) technology in web-based retrieval of *CMNet's* Emperor images using the SIMPLIcity technology he developed when he was at Stanford (Figure 6). As the collaborator to Wang's NSF/ITR (Information Technology Research) project, we in turn have provided him with the large number of attractive images of the Emperor's terracotta warriors and horses as well as other images of similar patterns and colors, together with extensive metadata information. These have proven to be invaluable to him because these images and data helped him to further refine his CBIR techniques [Chen 2002; Wang 2002; Wang 2003]. This kind of collaboration has been a good cross-fertilization example.

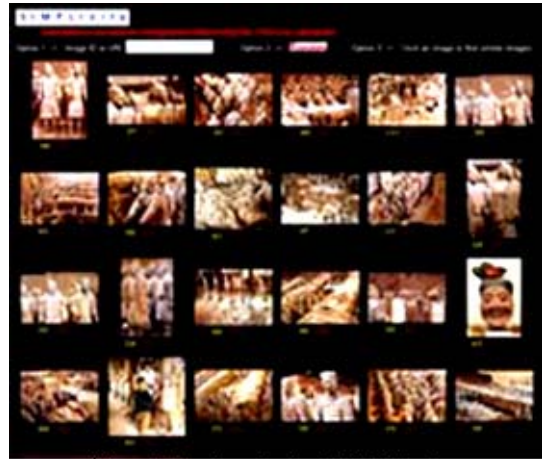


Fig. 6. Retrieval using SIMPLIcity

The SIMPLIcity CBIR has enriched the retrieval of *GMNet* beyond the usual traditional retrieval by metadata fields as the real-time online demonstration will show. Additional few



Fig.7. Random images show when user selects the image database

figures are shown here to demonstrate how this kind of retrieval technique can help to stretch the innovative use of a large image resource collection. Figure 7 shows that when a user enters the Emperor Image Database, images will show up randomly. The user can choose to search either by traditional retrieval method for precise images -- by title, keyword, location, date, description etc. or by CBIR. If latter, for example, on can ask the system to show all the images matching the first image on the upper left, then almost instantly, images of similar color or pattern will show up (Figure 8). On that

screen, descriptive information of any chosen image can be provided if asked or several derivatives of the image can also zoomed as shown in Figure 9.



Fig. 8 (left). Images of the same color, shape and pattern



Fig. 9. Chosen image is further enlarged

To overcome the problem related to “copyright” and intellectual property issues, all images larger than the thumbnail size would be displayed with a dynamically generated digital watermark of the image owner. This has facilitated the contribution of image content resources to *GMNet*.

3.4. Emperor’s Digital Video and Informedia Technologies

Carnegie Mellon University’s well-known Informedia Project is one of the six original NSF/DLI-1 projects. It has continued its further development in digital video related technologies and tools ever since 1995. Collaboration between Informedia and *CMNet* has enhanced perspectives from cultural and historical video documentaries. Its multi-lingual (English and Chinese) has also posed challenges in its speech recognition research (Wactlar 2002). When the Informedia technology is ready for web-based use, *CMNet* will be ready to explore it. Figure 10 shows some of the screens generated from the latest collaboration (more real-time demo will be given). Upper left shows that when “emperor” is searched, 60 video segments with that word have been identified and retrieved, these can be visualized in timeline as shown in the lower left. Map is shown in the upper right, and when one of the video is chosen, the video will play in the upper right of the lower right screen, and below that, the actual text will also be displayed with the word “emperor” highlighted in red.

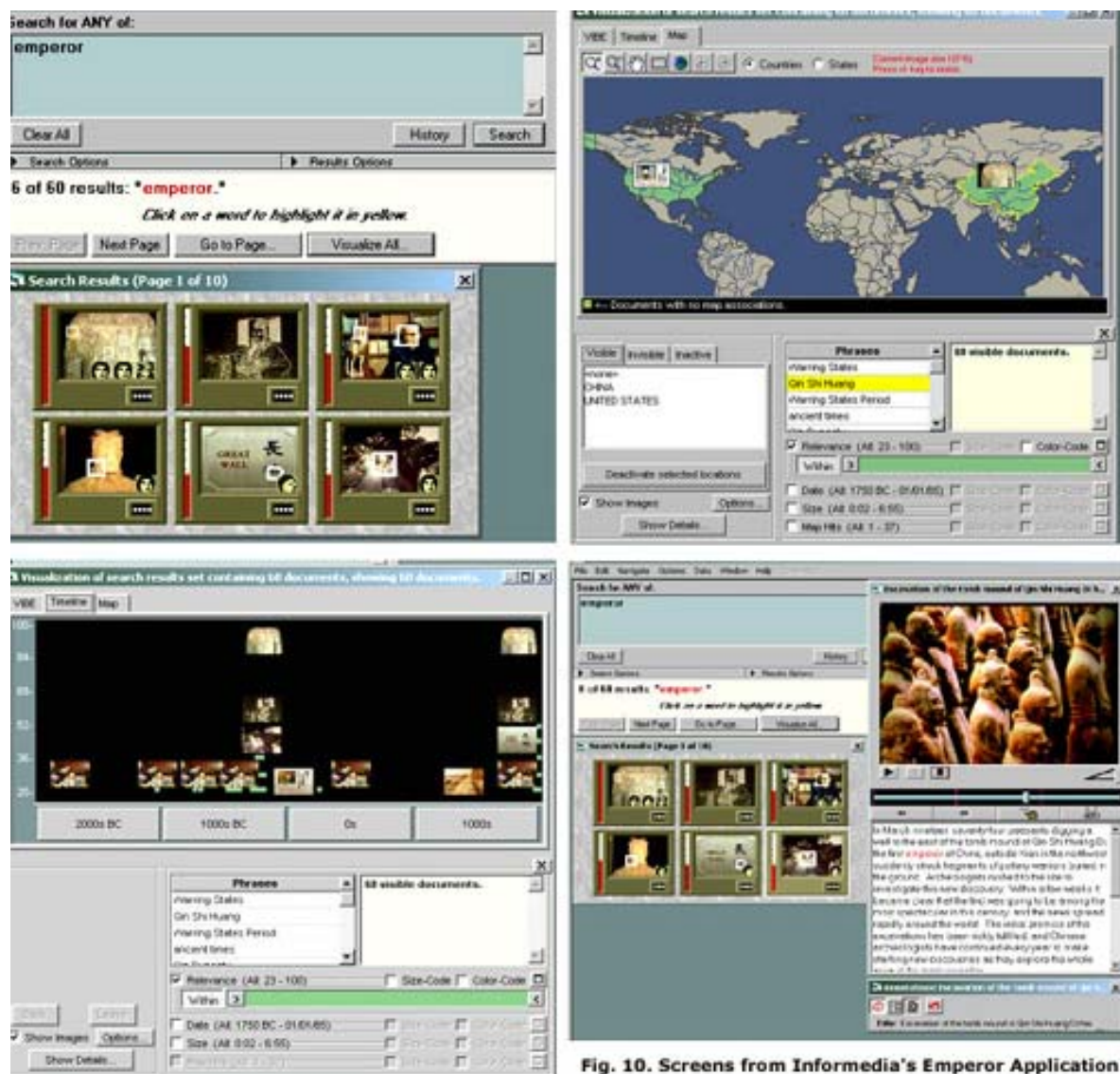


Fig. 10. Screens from Informedia's Emperor Application

3.5. Collaborative Project Restore with University of Florence

Prof. Piero Baglioni and his research group at the Center for Colloid and Interface Science (CSGI) of the University of Florence have developed an effective and sophisticated nanoparticle technology for restoring damaged or degraded cultural relics. Detailed information on this nanoparticle chemistry technology in restoration is summarized in several recent articles [Baglioni 2003]. What is worthwhile to mention and demonstrated here is that the invaluable restoration images of the Baglioni group in Florence are being organized and presented in *GMNet*, known as *Project Restore*. This multimedia digital collection created during the restoration process (Baglioni 2003, Figure 11), was one of the first to extend the



Fig. 11. Retrieved Pre- & Post-restoration Images

CMNet's scope and activities to *GMNet*. Figure 11 shows the pre- and post- restoration images retrieved from the *Project Restore*, it is clear that these images together with the associated descriptive information are of substantial significance to art historian and art specialist and students. The information on the restoration techniques and processes, to be included in the Project Archive of the *GMNet* is also of great value to preservation and restoration specialists as well as chemists.

4. GMNET'S DIGITAL MEDIA DEVELOPMENT AGENDA

In the last year, *GMNet* (Figure 12) has grown fast in digital image community building and in subject content expansion as shown in Figure 3. It has successfully established a model for international collaboration and sharing in both content development as well as cross-disciplinary R&D. Once it is possible to develop a multimedia digital library in one subject disciplinary or for one geographical area, it can be upward scalable to include more subject areas and bigger geographical areas from country to continent and to the world.



Figure 12. The Home Page of GMNet

It is fair to state that the current development has mainly been focused in the development of a functional global “still images” digital library on cultural heritage, although a small amount of digital video contents are included. Much more efforts will be made to effectively handle other multimedia format, specifically digital videos, sounds from simple audio to music, and 3-D images. These will require the collaboration with a new group of technologists in these specialized areas. I am, as usual, ready to explore possibilities with any specialists with these specialized capabilities, who also share the same interest and vision of *GMNet* in developing a effective global multimedia digital library on cultural heritage of the world.

5. CONCLUSION

In line with the vision stated in the US President's Information Technology Advisory Committee's *Digital Library Panel Report* [US PITAC, 2001] and the recommendations of *the Report of the DELOS-NSF Working Group on Digital Imagery for Significant Cultural and Historical Materials* [Chen 2002a], I would like to engage in more digital library community building, solicit more content providers to contribute their valuable resources, and increase collaborative activities in many essential fronts – content development and management on one hand, and cutting-edge technology development on the other. As I stated in my talk at the *ICDL* meeting a few days ago in Delhi [Chen, 2004], this is an exciting time for global digital library development. India has been in the forefront of developing a national digital library portal. With its history and incredibly rich cultural heritage, there are so much India can share with the rest of the world. From the technological development point of view, I am at awe at the capabilities of computer science specialists in Bangalore. I hope that my lecture has provided more examples to foster more fruitful collaboration between technology and humanities.

REFERENCES

- (Baglioni 2003) Baglioni, P., R. Giorgi and C. C. Chen, "Nanoparticle technology saves cultural relics, and potential for a multimedia digital library," *DELOS/NSF Workshop on Multimedia Contents in Digital Libraries*, Crete, Greece, June 2-3, 2003.
- (Chen 1993) Chen, C. C., "Technological potentials for the Global Digital Library," *Proceedings of the International Conference on National Libraries – Towards the 21st Century*, National Central Library, Taipei, Taiwan, April 20-24, 1993. Taipei, Taiwan: National Central Library, 1993. pp. 859-868.
- (Chen 2001) Chen, C. C., ed. *Global Digital Library in the New Millennium: Fertile Ground for Distributed Cross-Disciplinary Collaboration*. Beijing, China: Tsinghua University Press, 2001.
- (Chen 2001a) Chen, C. C., "Chinese Memory Net (CMNet): A model for collaborative global digital library development," *Global Digital Library in the New Millennium: Fertile Ground for Distributed Cross-Disciplinary Collaboration*. C. C. Chen, ed. Beijing, China: Tsinghua University Press, 2001. pp. 21-32.
- (Chen 2002) Chen, C. C. and J. Z. Wang, "Large-scale emperor digital library and semantics-sensitive region-based retrieval," *Proceedings of the Digital Library - IT Opportunities and Challenges in the New Millennium: An International Conference*, Beijing, July 9-11, 2002. pp. 454-462.
- (Chen 2002a) Chen, C. C. and K. Kiernan, eds. *Report of the DELOS-NSF Working Group on Digital Imagery for Significant Cultural and Historical Materials*. December 2002.
http://dli2.nsf.gov/internationalprojects/working_group_reports/digital_imagery.html
- (Chen 2003) Chen, Ching-chih, "Designing multimedia digital libraries with content development and management: Technology potential and challenges," in *Online Proceedings of DELOS/NSF Workshop on Multimedia Contents in Digital Libraries*, Crete, Greece, June 2-3, 2003.
- (Chen 2004) Chen, Ching-chih, "Global Memory Net: Potential and challenges for archiving and sharing cultural and heritage resources," an invited keynote presented at the *International Conference on Digital Libraries*, Delhi, February 26, 2004.
- (Soo 2002) Soo, V. W., C. Y. Lee, C. C. Yeh, and C. C. Chen, "Using sharable ontology to retrieve historical images," accepted for presentation at the *ACM/IEEE Joint Conference of Digital Libraries*, Portland, OR, July 15-18, 2002. *Proceedings of the ACM/IEEE JCDL Conference, 2002*. pp. 197-198.
- (Soo 2003) Soo, V. W., C. Y. Lee, C. C. Lin, S. L. Chen, and C. C. Chen, "Automated semantic annotation and retrieval based on sharable ontology and case-based learning techniques," full-length paper accepted for present

at the ACM/IEEE Joint Conference of Digital Libraries, Houston, TX, May 29, 2003. *Proceedings of the ACM/IEEE JCDL Conference*, 2003.

(US PITAC 2001) President's Information Technology Advisory Committee. Panel on Digital Libraries. Report to the President. *Digital Libraries: Universal Access to Human Knowledge*. Washington, DC, February 2001.

(Wactlar 2002) Wactlar, H. D. and C. C. Chen, "Enhanced perspectives for historical and cultural documentaries using Informedia technologies," *Proceedings of the ACM/IEEE Joint Conference of Digital Libraries*, Portland, OR, July 15-18, 2002. pp. 338 – 339.

(Wang 2002) Wang, J. Z., J. Li and C. C. Chen, "Interdisciplinary research to advance digital imagery indexing and retrieval technologies for Asian art and cultural heritages," *Proc. ACM Multimedia, Workshop on Multimedia Information Retrieval*, Juan Les Pins, France, December 2002. 6 pp.

(Wang 2003) Wang, J. Z., J. Li and C. C. Chen, "Machine annotation for digital imagery of historical materials using the ALIP system," *DELOS/NSF Workshop on Multimedia Contents in Digital Libraries*, Crete, Greece, June 2-3, 2003.

About Prof. Ching-chih Chen

Dr. Ching-chih Chen is Professor of the Graduate School of Library and Information Science, Simmons College, Boston, USA. A sought-after international consultant and speaker in over 40 countries, she is an author and editor of over 35 books and more than 150 scholarly journal articles. She was the Chief Conference Organization of a series of 12 *International Conferences on New Information Technology (NIT)* from 1987-2001 in different parts of the world.

Currently she directs a major US National Science Foundation/International Digital Library Project, *Chinese Memory Net*. Using that as a model for archiving, content building in specifically image and video areas, organization and retrieval, and international collaboration in technology related R&D activities, the project has expanded to "*Global Memory Net*" ready to include valuable cultural and historical resources of the world.

Served on the US President's Information Technology Advisory Committee from 1998-2002 with Prof. Raj Reddy of CMU, she is also the Co-PI of the *China-US Million Book Digital Library Project* with Prof. Raj Reddy. Active in the Digital Library area, she is the co-Chair of the *4th ACM/IEEE Joint Conference on Digital Libraries (JCDL)* of 2004 to be held in Tucson, Arizona in early June 2004. She is on the Advisory Board of DELOS (the European Digital Network of Excellence), US Co-Chair of the NSF/DELOS Working Group in Digital Imagery for Significant Cultural, Historical and Heritage Materials, and the Co-editor for the *Journal of Digital Library's* Special issue on *Multimedia Contents in Digital Libraries*.