LARGE-SCALE EMPEROR DIGITAL LIBRARY AND SEMANTICS-SENSITIVE REGION-BASED RETRIEVAL

Ching-chih Chen
Graduate School of Library and Information Science
Simmons College
Boston, MA 02115 USA
+1 617 521 2804
chen@simmons.edu

James Z. Wang
School of Information Sciences and Technology
The Pennsylvania State University
University Park, PA 16802 USA
+1 814 865 7889
jwang@ist.psu.edu

ABSTRACT

In recent years, with the advent of fast-speed, broadband telecommunications networks, and information technology research for digital libraries, more and more distributed digital libraries with multimedia information have been developed everywhere in the world. While technologies are available, there is insufficient large-scale and coordinated digital content development. Furthermore, state-of-the-art technologies developed in the research labs are rarely used to assist the content development and content analysis. This paper first discusses how the NSF/IDLP Project, *Chinese Memory Net* has capitalized on the rich multimedia resources in both analog and digital formats of an earlier cultural documentaries products, the award-winning *The First Emperor of China* videodisc and multimedia CD, to further build large-scale digital contents of significant museum and historical/cultural/heritage materials for productive international collaboration among experts from interdisciplinary fields. The content has been used as a testbed for technology research in the area of semantics-sensitive region-based image retrieval. We demonstrate the use of the SIMPLIcity technology in browsing and retrieving of images.

Keywords

Large-scale content development, Humanities computing, International Collaboration, Historical/Cultural Materials, image retrieval, region-based, wavelet

INTRODUCTION

As early as in 1984, when the use of the Internet, CD-ROMs, interactive videodisc and multimedia technologies were either at its beginning or still unknown, a set of two double-sided videodiscs on The First Emperor of China and his 7000 fascinating terracotta figures of warriors and horses in the ancient capital of China, Xian, were created by PROJECT EMPEROR-I with a major funding from the US National Endowment for the Humanities' Library in Humanities Program. In 1991 and 1994, popular versions of award-winning interactive videodisc and multimedia CD-ROM products were produced and published by the Voyager Company [1]. As the producer, the PI - Ching-chih Chen - and her research and production teams took great care in gathering, assembling and creating multimedia materials attempting to meet the highest scholarship possible. Great efforts were made to communicate messages and create courseware with highly selective sequences of still images and video of artifacts and locations, audio narrations in both English and Chinese, expert commentaries, interviews, and music, description texts, and linking reference tools and references.

The experience and knowledge gained from these multimedia R&D efforts led to her further promotion of the creation of digital knowledge base [2], and the concept of "global digital library" [3]; and contributed largely to the successful proposal of *Chinese Memory Net (CMNet)* to the NSF/IDLP Program [3].

As potential multi-national collaborators of *CMNet* began to discuss possible and challenging collaborative research, it became clear that there is a great need for large-scale, multi-format, multi-modal, multi-lingual content building. Our earlier experience in creating multimedia applications convinced us that "humanities" content, while having its special unique features, could use similar technologies and techniques. Thus, it is reasonable to expect the technologies, techniques, and tools developed thus far for digital libraries can be applied to the effective use of significant historical, cultural and heritage materials as well. Since "intellectual property" issues related to large-scale content building are so paramount for any digital library creations, it would seem natural for *CMNet* to capitalize on the PI's earlier cultural documentaries and very rich raw information resources, which include:

- A set of two double-sided videodiscs with 108 analog images for each disc and a one-sided popular video disc [1];
- A digital multimedia CD with several-hour interactive courseware [1];
- Numerous hours of analog tapes (oral history materials) only small part of them are included in the videodiscs;
- Four Photo-CDs of multi-level resolutions;
- Thousands of slides of the most significant still images with keyword-based descriptive information of these images;
- Several analog films most of which have not been utilized fully;
- English and Chinese audios for all the videos and oral history materials on the videodiscs,
- Many others.

CONTENT BUILDING

CMNet has made extensive efforts in several directions. Limited space permits the mention only a few examples:

1. To create over 3000 high-resolution images from the original slides (not videodiscs due to its low TV-level resolution). Each image is scanned at 1200 dpi and above with image size over mostly over 10MB to 75 MB for each image in its compressed format. In addition, 5 additional derivatives -- 1536, 768, 384, 192, and 96 pixels for the widest side (either vertical or horizontal) were created for each image with the smallest one serving as an icon. This permits the zooming effect of each image object. A prototype demonstration exhibit has been mounted by using the Insight of Luna Imaging, Inc. in Culver City, CA (Figure 1).

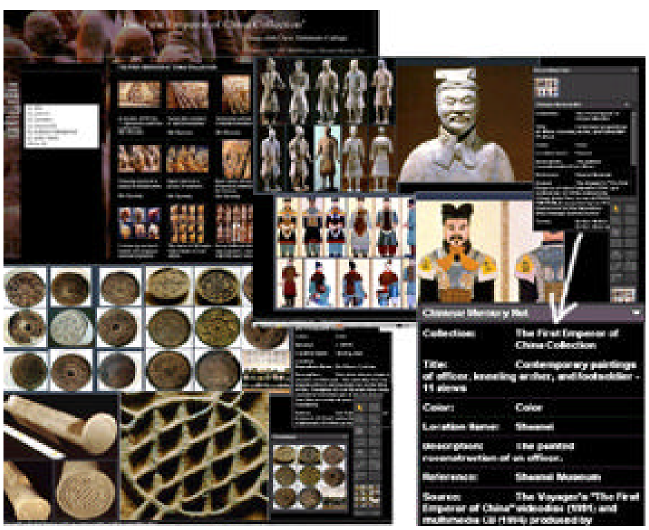

Figure 1. Selected images of Emperor Image Exhibit Collection (3 Screens) with Enlarged Metadata Info

The metadata information here can be crossed walked from Emperor's own to other standard formats, such as Dublin Core, CIMI, VRA1.0, VRA3.0, US Marc, etc. More research will be needed to study this aspect.

2. To create comprehensive image database information for each of the 3000+ significant images with extensive metadata, description, right and source information, as well as references etc. as shown in Figure 2. Every single field can be quickly searched (Figure 3).

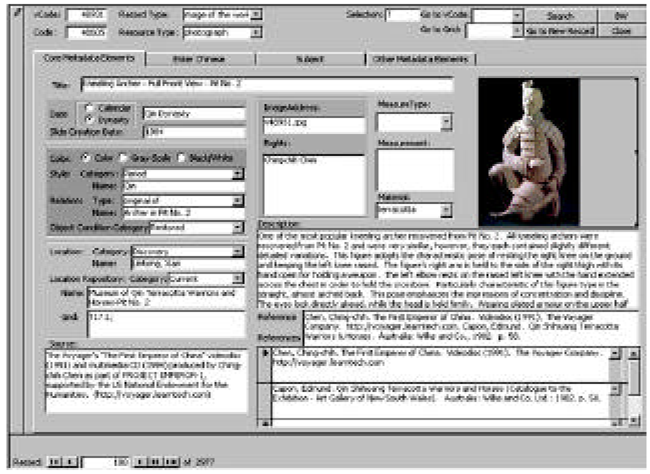

Figure 2. A Part of A Typical Image Database Record

POTENTIALS FOR INTERNATIONAL COLLABORATION

The few examples given here are only the very tip of a big iceberg. It is clear that content building is extremely labor-intensive. Yet, once the comprehensive knowledge base is created, the reward is great. It opens up incredible opportunities for international collaboration. Among a dozen possibilities related to multiform, multi-modal, multilingual, metadata R&D activities, this paper begins to show, for example, Figure 1 suggests perfect data for content-base image retrieval -- by shape, color, semantic context, etc. [4]. The digital videos and their associated data offer great possibilities for using Informedia technologies for enhanced perspectives [5]. The rich metadata and descriptive information (Figures 2-3) offers great potential for collaborative metadata cross-mapping research as well as research in using sharable ontology [6].

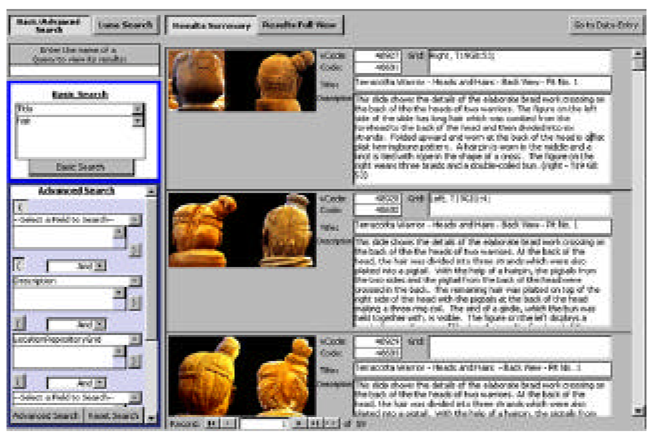

Figure 3. All Images Relevant to "Hair" Are Retrieved

IMAGE RETRIEVAL WITH THE SIMPLIcity SYSTEM

Many content-based image retrieval (CBIR) systems have been developed since the early 1990s. Examples include the IBM QBIC System [7] developed at the IBM Almaden Research Center, the Photobook System developed by the MIT Media Lab [8], the VIRAGE System [9] developed by the Virage Incorporation, the VisualSeek System [10] developed at Columbia University, the systems [11] developed at U.C. Santa Barbara, the WBIIS and SIMPLIcity systems [4] developed at Stanford University, the Blobworld System [12] developed at U.C. Berkeley.

Typically, CBIR systems extract a signature for every picture based on its pixel values, and define a measure or distance for comparing images. The signature serves as an image representation in the `view' of a CBIR system. The components of the signature are called features. One advantage of a signature over the original pixel values is the significant compression of image representation. However, a more important reason for using the signature is to gain improved correlation between image representation and semantics. Actually, the main task of designing a signature is to bridge the gap between image semantics and the pixel representation, that is, to create a better correlation with image semantics.

Existing general-purpose CBIR systems roughly fall into three categories depending on the approach to extract signatures: histogram, color layout, and region-based search. There are also systems that combine retrieval results from individual algorithms by a weighted sum matching metric or other merging schemes.

There is a growing trend in the field of image retrieval to develop efficient region-based systems. A region-based retrieval system applies image segmentation to decompose an image into regions, which correspond to objects if the decomposition is ideal. The object-level representation is intended to be close to the perception of the human visual system (HVS). However, image segmentation is nearly as difficult as image understanding because the images are 2-D projections of 3-D objects and computers are not trained in the 3-D world the way human beings are.

Two of the earlier region-based systems, the NeTra and the Blobworld, compare images based on individual regions. Although querying based on a limited number of regions is allowed, the query is performed by merging single-region query results. The motivation is to shift part of the comparison task to the users. To query an image, a user is provided with the segmented regions of the image, and is required to select the regions to be matched and also attributes, e.g., color and texture, of the regions to be used for evaluating similarity. Such querying systems provide more control to the user. However, the user's semantic understanding of an image is at a higher level than the region representation. For objects without discerning attributes, such as special texture, it is not obvious for the user how to select a query from the large variety of choices. Thus, such a querying scheme may add burdens on users without significant reward. On the other hand, because of the great difficulty of achieving accurate segmentation, systems in often partition one object into several regions with none of them being representative for the object, especially for images without distinctive objects and scenes.

Figure 4. An image and its 3-level wavelet transform

In our SIMPLIcity project, originally developed at Stanford University, we attempt to develop a region-based image retrieval system that is not relying heavily on the correct image segmentation. We developed a fast statistical clustering based image segmentation algorithm that is able to segment an image with 512 x 512 pixels in less than a second on a Pentium PC. We define a robust region-based similarity measure, the Integrated Region Matching (IRM) metric. It incorporates the properties of all the segmented regions so that information about an image can be fully used to gain robustness against inaccurate segmentation. Image segmentation is an extremely difficult process and is still an open problem in computer vision.

To segment an image, SIMPLIcity partitions the image into blocks of pixels and extracts a feature vector for each block. The feature vector is computed in the wavelet transform domain in the LUV color space. Figure 4 shows an image with its 3-level wavelet transform. The k-means algorithm is used to cluster the feature vectors into several classes with every class corresponding to one region in the segmented image. Since the block size is small and boundary blockyness has little effect on retrieval, we choose block-wise segmentation rather than pixel-wise segmentation to lower computational cost significantly.

Six features are used for segmentation. Three of them are the average color components in the block. The other three represent energy in high frequency bands of wavelet transforms, that is, the square root of the second order moment of wavelet coefficients in high frequency bands.

After the region segmentation, we extract a feature vector for each of the regions in the image to represent the information stored in the region. This feature vector captures the color, texture, and shape of the region. Depending on the semantic complexity, images can be segmented into different numbers of regions.

Traditionally, region-based matching is performed on individual regions. The IRM metric we have developed has the following major advantages:

- Compared with retrieval based on individual regions, the overall ``soft similarity" approach in IRM reduces the adverse effect of inaccurate segmentation, an important property lacked by previous systems.
- In many cases, knowing that one object usually appears with another helps to clarify the semantics of a particular region.
- By defining an overall image-to-image similarity measure, the SIMPLIcity system provides users with a simple querying interface. To complete a query, a user only needs to specify the query image.

Another important feature of the SIMPLIcity system is its sensitivity to semantics. Semantic classification methods are used to categorize images so that semantically-adaptive searching methods applicable to each category can be applied. At the same time, the system can narrow down the searching range to a subset of the original database to facilitate fast retrieval.

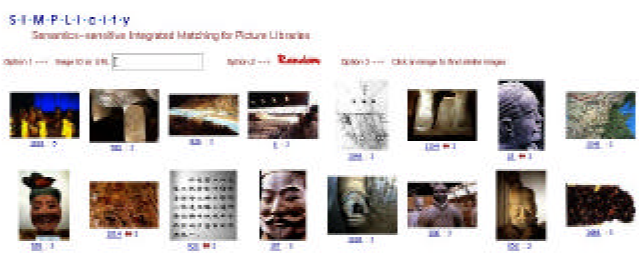

Figure 5. Random browsing of the Emperor image database using the SIMPLIcity system.

The SIMPLIcity system was developed on UNIX platforms using the C programming language. We tested the system on a Pentium PC with a 800 MHz CPU. It takes less than a second for the computer to index each image in this database of about 2000 images. A query takes less than 0.02 second of CPU time. The system does not require any textual annotations in the matching process.

The system can be used in the following ways:

- Random browsing of images stored in the Emperor database (Figure 5).
- Given an image in the database, search for visually similar pictures. (Figure 6)
- By providing an URL of an image on the Web, search for related images in the Emperor database (Figure 7).
- Search the database using a hand-drawn sketch.

(A)

(B)

Figure 6. Two similarity search examples. The upper-left corner of each block of images is the query image. The images are ranked according to their similarities to the query image.

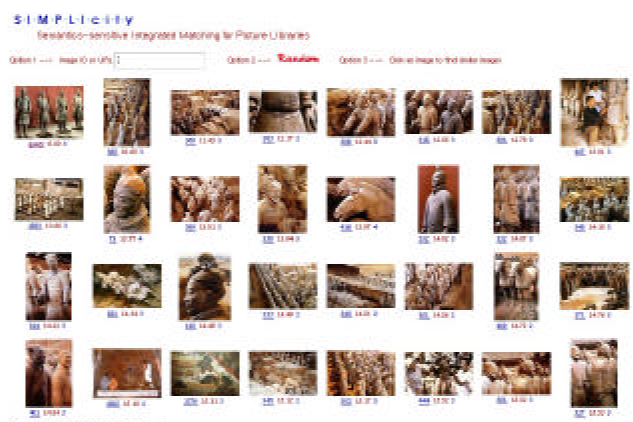

Figure 7. An image from the Web, upper-left corner image, is used as the query to find related images in the Emperor database. The image was downloaded from http://www.unc.edu/courses/hist033

CONCLUSIONS

While the content building efforts have already begun to yield productive results, it is fair to conclude that much more can be expected. When *CMNet* is teaming up with the digital contents created by the *China-US Million Book Digital Library Project*, the prospect is even more exciting. The SIMPLIcity image retrieval project is currently developing statistical modeling methods for automatic image annotation.

ACKNOWLEDGMENTS

This paper is based on work supported by the National Science Foundation's IDLP Project, *Chinese Memory Net*, under Grant No. IIS-9905833. The SIMPLIcity work was supported in part by the NSF DL2 Initiative under Grant No. IIS-9817511, SUN Microsystems, and the PNC Foundation.

REFERENCES

1. Chen, Ching-chih. *First Emperor of China*. The Voyager Company. Interactive Videodisc, 1991. Multimedia CD, 1994, http://www.voyagerco.com/
2. Chen, Ching-chih. Multimedia and the First Emperor of China: Moving Toward A Digital Knowledge Base. *Multimedia Today* (IBM), 2, 2 (April 1994), 68-71.
3. Chen, Ching-chih. *Chinese Memory Net (CMNet):* A Model for Collaborative Global Digital Library Develop-ment, in *Global Digital Library Development in the New Millennium: Fertile Ground for Distributed Cross-Disciplinary Collaboration*. Edited by Ching-chih Chen. Beijing: Tsinghua University Press, 2001. pp. 21-32.
4. Wang, James Z. Integrated Region-based Image Retrieval. Dordrecht: Kluwer Academic Publishers, 2001.
5. Wactlar, Howard D. and Ching-chih Chen. Enhanced Perspectives for Historical and Cultural Documentaries Using Informedia Technologies. Submitted for JCDL, 2002.
6. Soo, Von-Wun, Chen-Yu Lee, Chao-Chun Yeh, and Ching-chih Chen. Using Sharable Ontology to Retrieve Historical Images. Submitted for JCDL, 2002.
7. Flickner, M, H. Sawhney, W. Niblack, J. Ashley, Q. Huang, B. Dom, et al. Query by image and video content: The QBIC system. *IEEE Computer*. vol. 28, no. 9, 1995.
8. Pentland A., R. W. Picard, S. Sclaroff. Photobook: tools for content-based manipulation of image databases. *Proc. SPIE*. vol. 2185, pp. 34-47, February 7-8, 1994.
9. Gupta A., R. Jain. Visual information retrieval. *Comm. ACM*, vol. 40, no. 5, pp. 70-79, May 1997.
10. Smith J. R., S.-F. Chang. VisualSEEk: a fully automated content-based image query system. *Proc. ACM Multimedia*. pp. 87-98, Boston, MA, Nov. 1996.
11. Ma W. Y., B. Manjunath. NaTra: A toolbox for navigating large image databases. *Proc. IEEE Int. Conf. Image Processing*, pp. 568-571, Santa Barbara, 1997.
12. Carson C., M. Thomas, S. Belongie, J. M. Hellerstein, J. Malik. Blobworld: a system for region-based image indexing and retrieval. *Proc. Visual Information Systems, LNCS*. vol. 1614, pp. 509-516, Springer, Amsterdam, Jun. 1999.