



*Qin terra cotta figures of warriors and horses from PROJECT EMPEROR-I.*

# Large-Scale Image Processing

by Ching-chih Chen

Until recently, most of the available machine-readable databases dealt with textual and numeric information and/or data. Some may argue that words and numbers are coded on forms which gen-

erate relationships that give the data meaning and, thus, that they actually provide "pictures" of data. However, these "pictures" certainly do not represent the complexity of real images in our world.

We are all familiar with the saying that "a picture is worth a thousand words." As a rule, a picture is not word-based and may likely contain many complex

concepts. Pictures can mean different things to different people, depending on an individual's view of an image in a given context. This is why most of the existing manual image/picture indexing and cataloging methods/schemes, most of which use keyword approaches, can serve only as location guides, rather than as adequate and effective retrieval tools for non-numeric or textual data.

*Ching-chih Chen is professor and associate dean at the Graduate School of Library and Information Science at Simmons College in Boston.*

Large-scale image processing and management represents an exciting challenge for information retrieval. PROJECT EMPEROR-I meets the challenge with the application of interactive videodisc to present and interpret a major historical period of China's past and one of the most significant archaeological finds of this century.

### PROJECT EMPEROR-I

Among the products of PROJECT EMPEROR-I is a set of two-sided videodiscs entitled "The First Emperor of China." The first disc contains 108,000 frames of analog images, many of which are multi-color pictures of Qin artifacts and the magnificent terra cotta figures and pits. While many of these images deal with the same topics, they present very different visual pictures. Using the terra cotta figures as an illustration, one can imagine thousands of pictures, each with a variation from the others, but all representing warriors and horses. So, in any given catalog, one might find the following pictures:

- one single complete warrior or horse;
- numerous full-size warriors or horses;
- combinations of types of warriors and horses;
- numerous warriors and horses in the real context of the site;

- close-up views of figures.

Of course, this list could go on forever, but it is clear that keyword approaches to the organization of these images, using words such as *horse*, *warrior*, *Pit #1*, etc., without adequate and detailed descriptions and modifiers, are too oversimplified to yield relevant and precise retrieval of needed images.

### PROJECT EMPEROR-I's Image Database System Configuration

The PROJECT EMPEROR-I image database system demonstrates the application of computer processes and videodisc technology in a unique way. It pairs the computer's textual display and search/retrieval capability with videodisc technology's high density of image storage and high-speed random access and retrieval of images. The system's configuration consists of the following:

- Hardware—IBM/XT, AT or compatible; color monitor, printer, videodisc player;
- Software—picture/image cataloging and retrieval software program for image control and index, specifically Image Concept's C-Quest;
- Image Base—more than 5000 still-frame slides and related video stored on videodisc;

- Textual database—constructed on microcomputer using C-Quest.

### Cataloging and Retrieval Software

Because of the importance of the software to a successful image processing project, the selection of C-Quest was a major step. While several other picture cataloging and retrieval software packages were considered for PROJECT EMPEROR-I, C-Quest was selected on the basis of its features.

C-Quest eliminates keywords as an approach to image organization. As stated previously, this is critical to precise retrieval.

The software has a flexible synonym capability, enabling system users to apply any terms and descriptions familiar to them as searchable terms. C-Quest is also a knowledge-based filing system governed by hierarchically-structured synonym thesauruses, allowing users to search subject matter using familiar terms.

With the PROJECT EMPEROR-I software, each image can be cataloged and accessed with as many as twenty terms and each term can also include modifiers. For example, to catalog the photograph accompany this article, one can use the following descriptors:

- Warrior—close-up
- Head—front
- Hand—side view
- Pit #1—interior

The image records consist of many retrievable and searchable fields. Once entered, searches can be performed on any one or combination of the fields.

Among other key features of the PROJECT EMPEROR-I software are its relative simplicity with little training needed to access the image database; ability to restructure picture files; index counter for estimating search time; and "cut and paste" for dictionary modification.

### Development Continues

PROJECT EMPEROR-I is still at the developmental stage, experimenting with C-Quest and testing the various features. While we will undoubtedly suggest areas for further development and improvement, PROJECT EMPEROR-I demonstrates that there are numerous complex issues and problems involved in large-scale image processing, organization and management. These issues deserve close attention.

## Typical Record Entry Form PROJECT EMPEROR-I

UNIQUE NUMBER .....	000002
FRAME ID .....	
DISCOVERY SITE . . . . .	SHAANXI
CURRENT LOCATION . . . . .	
INFORMATION SOURCE . . . . .	
PUBLICATION SOURCE . . . . .	
DATE MM/DD/YY .....	/ /
FORMAT .....	<H_V_S> Horizontal
FILMTYPE .....	<T_C_B_X_I> Color
DISK NUMBER & SIDE .....	1 1
STARTING FRAME NUMBER .....	48796
ENDING FRAME NUMBER .....	

# Optical Disk Technology

Current Trends in the Optical  
Storage Industry

Large-Scale Image Processing  
and Management

Creating a CD-ROM with a  
Simulator

Videodiscs and Training

CD-I: A Powerful Interactive  
Audio/Video System

Data Structures for CD-ROM

Standards Issues for Optical  
Publishing

Challenges for Federal  
Information Policy

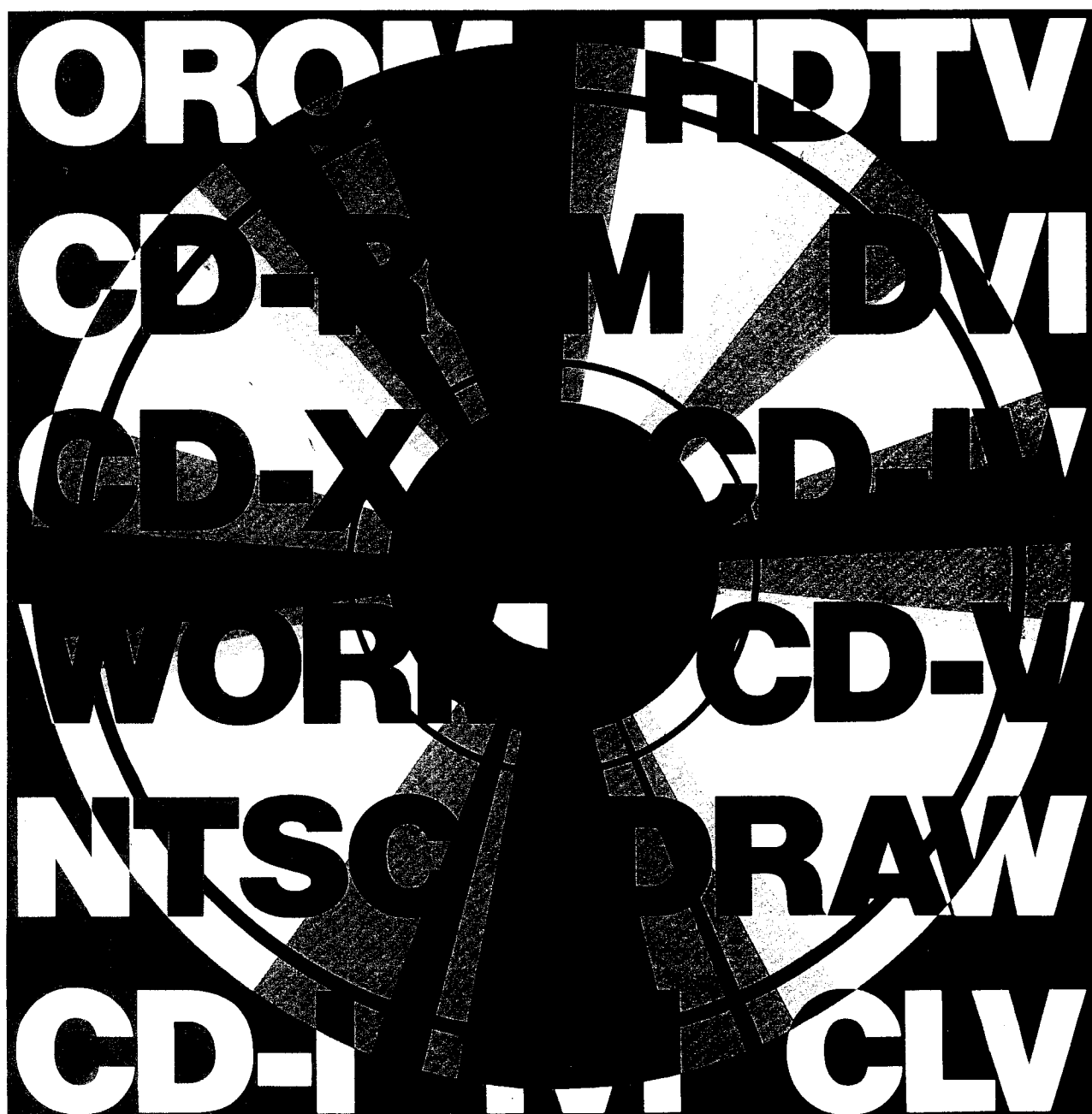
# BULLETIN

of the American Society for Information Science

August/September 1987

Volume 13, Number 6

ISSN: 0095-4403 CODEN: BASICR



**Editor and Publisher**  
Linda Resnik

**Associate Editor**  
Stephen E. Morrison

**Guest Editor**  
George L. Abbott

**Advisory Board**  
Rep. George E. Brown, Jr.  
Robert L. Chartrand  
Kathleen Criner  
Margaret T. Fischer  
Thomas D. Harnish  
Donald W. King  
Joseph H. Kuney  
Richard E. Wiley  
Donald H. Kraft

The American Society for Information Science (ASIS) is a non-profit professional association organized for scientific, literary and educational purposes and dedicated to the creation, organization, dissemination and application of knowledge concerning information and its transfer.

The official ASIS journal is the *Journal of the American Society for Information Science*, published bi-monthly for the Society by John Wiley & Sons, Inc., 605 Third Ave., New York, NY 10158. The *Journal* contains scholarly papers in the fields of information science and technology.

The *Bulletin of the American Society for Information Science* is a news magazine concentrating on issues affecting the information field; management reports; opinion; and news of people and events in ASIS and the information community.

Opinions expressed by contributors to ASIS publications do not necessarily reflect the position of official policy of either their employers or the Society.

---

#### ASIS Officers

Thomas H. Hogan, President  
Martha E. Williams, President-elect  
Julie A.C. Virgo, Past President  
N. Bernard Basch, Treasurer

#### Directors

Trudi Bellardo  
Carol Wasserman Diener  
Marjorie Hlava  
Edward Kazlauskas  
Lois F. Lunin  
Ann E. Prentice  
G. Daniel Robbins  
Gerard Salton  
Roy Tally (deputy)  
Mickie Voges (deputy)

#### Executive Director

Linda Resnik

# BULLETIN

of the American Society for Information Science Vol. 13, Number 6  
August/September 1987 ISSN 0095-4403 CODEN: BASICR

---

- Special Section**
- 11 Optical Disk Technology**  
George L. Abbott, Guest Editor
  - 12 Current Trends in the Optical Storage Industry**  
John C. Gale
  - 15 Large-Scale Image Processing**  
Ching-chih Chen
  - 18 Data Structures for CD-ROM**  
Hugh Marlor
  - 21 Creating a CD-ROM with the Help of a Simulator**  
Roy Tally
  - 22 Videodiscs and Training**  
Cathrine E. Snyder and Sheila L. Webster
  - 24 CD-I: A Powerful Interactive Audio/Video System**  
Bert Gall
  - 26 CD-ROM and Related Technologies: Challenges for Federal Information Policy**  
Karen J. Sy
  - 27 Standards Issues for Optical Publishing**  
Clifford A. Lynch

#### Opinion

- 36 The Information Age?**  
David Pannkuk

#### Departments

- 3 Newsfront**
- 6 Inside ASIS**
- 30 Information News**
- 32 Information Products**
- 35 Calendar**

---

The *Bulletin of the American Society for Information Science*, ISSN 0095-4403, is published bi-monthly, October through September, by the American Society for Information Science, 1424 16th Street, NW, Washington, DC 20036. POSTMASTER: Send address changes to the *Bulletin of the American Society for Information Science*, 1424 16th Street, NW, Washington, DC 20036. The *Bulletin of the American Society for Information Science* is entered for second-class postage paid at Washington, DC.

The subscription rate for ASIS members is \$11, which is included in the annual membership dues. Non-member subscriptions, and additional member subscriptions, are \$50 per year U.S., Canada and Mexico; \$60 per year other, postpaid in the U.S. Single copies and back issues of the *Bulletin* may be purchased for \$9 each. Claims for undelivered copies must be made no later than three months following the month of publication.

Where necessary, permission is granted by the copyright owner for libraries and others registered with the Copyright Clearance Center (CCC) to photocopy any copy herein for \$0.50 per page. Payments should be sent directly to CCC. Copying done for other than personal or internal reference use without the expressed permission of the American Society for Information Science is prohibited. Serial-fee code: 0095-4403/83 \$0 + .75

Copyright © 1987 by the American Society for Information Science.