Introduction to the Special Section on Digital Libraries: Representation and Retrieval

INTRODUCTION

"COMPUTER, skip ahead to the scene where they ride bicycles through the sky." "Are there other shots of this, maybe from a different viewpoint?" "What other fabrics do you have that look like this?" These are a few queries that might be asked of today's digital library systems. As consumers receive access to the world's collections of video, image, text, and other information, research is needed for tools that represent digital content and save people time retrieving what they want to find.

Although information retrieval is already a mature research area for text databases, its use for video and image libraries is relatively new. Some of the earliest and largest research efforts have been IBM Almaden's QBIC system [1], ISS's SWIM system [2], and MIT's Photobook and FourEyes systems [3], [4]. Pieces of these first systems have evolved into commercial products and can be explored interactively on the world-wide web [5], [6].

PAPERS IN THIS SPECIAL SECTION

Regular

Chen et al.

This special section was formed to publish novel and significant tools and techniques that facilitate access to the content of large data collections. Although a large percentage of the PAMI community focuses on visual pattern analysis and recognition, most existing digital libraries contain text, e.g., of technical paper abstracts, and can benefit from pattern analysis and intelligent matching algorithms.

The first paper, by Chen et al., addresses a problem in large-scale information retrieval, where vocabularies differ within databases and searches, and where formation of "concept spaces" using techniques of pattern analysis can improve retrieval performance.

Samet and Soffer

The remaining papers focus on visual content, starting with images of a symbolic nature. Samet and Soffer describe the MARCO map retrieval system, which performs analysis and classification converting map images to logical representations and allows users to identify symbols of interest from map legends and formulate queries such as "find campsites within one mile of a beach."

Ratha et al.

Ratha et al. present new results on a classical image database problem, that of fingerprint retrieval. Their method of indexing combines domain-specific and domain-independent features and is evaluated using a high-speed parallel implementation.

Sawhney and Ayer

The computational demands of content-based retrieval systems are significant, and perhaps greatest for digital video systems. Perhaps as a consequence, work on digital video libraries has not progressed as quickly as work on still images. Only one paper is included on digital video, that of Sawhney and Ayer. This paper makes no restrictions on the domain of the data in the digital library and directly addresses basic problems in representing dominant and multiple motions—key tasks in isolating camera and object motion for providing regions for content analysis.

Correspondence

Swets and Weng; Manjunath and Ma; Healey and Jain; Schneier and Abdel-Mottaleb

The remaining four papers involve image retrieval systems. Swets and Weng present an automatic method for finding discriminating features and apply it to retrieval of faces and other objects. Manjunath and Ma extract Gabor wavelet features and evaluate their performance on the Brodatz texture database, as well as on the browsing of aerial imagery. Healey and Jain combine several representations and algorithms to address the problem of retrieving and annotating multispectral satellite imagery. Schneier and Abdel-Mottaleb propose and evaluate an image indexing scheme based on JPEG compressed data.

Although the papers here all reflect the latest research in digital libraries, they represent only a few of the key issues. Other important research issues that are not addressed here include: systems that understand human behavior and action in video (the vast majority of video is about people), systems that can infer which features are most relevant for a search given subjective or time-varying examples from a user, and systems that learn how to improve their performance during tasks such as annotation, so that a user does not have to repetitively provide the same labels to the same data.

The field of digital libraries brings with it the challenge of a human in an interactive loop, working with the system

• *The authors are with the Media Laboratory, Massachusetts Institute of Technology, Cambridge, MA 02139.*
E-mail: {picard; sandyj}@media.mit.edu.

Manuscript received May 15, 1996.
For information on obtaining reprints of this article, please send e-mail to: transpami@computer.org, and reference IEEECS Log Number P96056.

and often changing the query as more data are revealed. Systems need to be developed that do not contain just one method of pattern analysis for one set of data, but that learn to combine multiple perceptual and semantic features and integrate common-sense and higher-level knowledge with low-level features for retrieval. The papers here also focus on visual and semantic information, where, in fact, there is also a demand for tools for audio retrieval (e.g., "any drum riffs like this?") as well as for other kinds of data such as in DNA matching or drug design.

Although tools that can "see" and "understand" the content of digital libraries are still in their infancy, they have reached the crucial point where they can provide substantial assistance to users in navigating through visual and nonvisual media. We look forward to continued research in this fast-moving and important new area.

A few words should be added about the logistics of this special section. We received 32 submissions, mostly from the vision and image analysis community, signifying a high response given the relative newness of the research area. All submissions were subjected to the standard rigorous PAMI review process. To maximize objectivity, we editors agreed to not submit our papers to this issue. Finally, papers had to demonstrate conclusions, preferably on large data sets; this is the point at which many papers containing fine ideas and preliminary results had to be rejected. Ninety-six reviews later, the papers presented here emerged as suitable for publication. A couple additional submissions have been directed to the regular PAMI pipeline for various reasons.

ACKNOWLEDGMENTS

We are grateful to all the authors who worked hard and contributed to this special section. Special thanks also to A.K. Jain, R. Kasturi, and S. Barrier who advised in many aspects of preparation for this special section. We are indebted to L. Ward who expertly handled all the correspondence and administration on the MIT end. We would like to express thanks to the following reviewers for their assistance: P. Aigrain, P. Beardsley, G. Benson, A. Bobick, A. Bovik, T. Cass, C.C. Chang, S.-F. Chang, R. Chellappa, H. Chen, L. Coetzee, D. Cooper, J. Dalton, L. Davis, A. Del Bimbo, D. Delanoy, S. Dickinson, M.-P. Dubuisson, M. Flickner, D. Forsyth, S. Gallant, K. Haase, A. Halaas, A. Hampapur, G. Healey, A. Heller, D. Jacobs, A. Jain, R. Jain, M. Johnston, T. Kanade, J. Kittler, G. Kopec, J. Kreyss, S.-Y. Lee, M. Lesk, F. Liu, Z.-Q. Liu, J. Malik, S. Mann, B.S. Manjunath, B. Moghaddam, M. Mohiuddin, F. Mokhtarian, A.D. Narasimhalu, C. Nastar, B. Neveitt, W. Niblack, D. Petkovic, C. Pinhanez, K. Popat, I. Rigoutsos, E. Riseman, R. Sacks-Davis, Y. Sato, E. Saund, H. Sawhney, S. Sclaroff, S. Shafer, K. Shearer, A. Sherstinsky, M. Shiffer, B. Smith, S. Smoliar, M. Swain, M. Szummer, A.M. Tekalp, J. Tenenbaum, J. Wachman, A. Yuille, R. Zabih, and H.-J. Zhang.

REFERENCES

[1] W. Niblack, R. Barber, W. Equitz, M. Flickner, E. Glasman, D. Petkovic, P. Yanker, C. Faloutsos, and G. Taubin, "The QBIC Project: Querying Images by Content Using Color, Texture, and Shape," W. Niblack, ed., *Storage and Retrieval for Image and Video Databases*, San Jose, Calif., pp. 173–181, SPIE, Feb. 1993.

[2] H.-J. Zhang, S.W. Smoliar, J.H. Wu, C.Y. Low, and A. Kankanhalli, "A Video Database System For Digital Libraries," ISS, Nat. Univ. Singapore, June 1994.

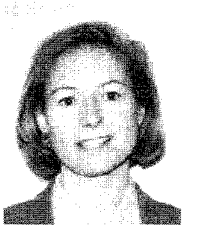
[3] A. Pentland, R. Picard, and S. Sclaroff, "Photobook: Tools for Content-Based Manipulation of Image Databases," *Int'l J. Computer Vision*, vol. 18, no. 3, 1996.

[4] T.P. Minka and R.W. Picard, "Interactive Learning Using a 'Society of Models,'" *Pattern Recognition, 1996*. To appear. Also appears as MIT Media Lab Perceptual Computing, TR #349.

[5] Virage Engine, 1995, http://www.virage.com/.

[6] IBM Ultimedia Manager System, http://wwwqbic.almaden.ibm.com/~qbic/qbic.html.

ROSALIND W. PICARD
ALEX PAUL PENTLAND
Guest Editors



Rosalind W. Picard earned a BEE from the Georgia Institute of Technology and was named a National Science Foundation Graduate Fellow in 1984. She earned the MS and ScD in electrical engineering and computer science from the Massachusetts Institute of Technology in 1986 and 1991, respectively. Dr. Picard was a Member of the Technical Staff at AT&T Bell Laboratories from 1984 to 1987, first in the Digital Signal Processing and Integrated Circuit Design group and later in the Visual Communications Research Department. In 1991 she joined the faculty at the M.I.T. Media Laboratory, in 1992 was awarded the NEC Development Chair in Computers and Communications, and in 1995 was promoted to associate professor. She has chaired several special sessions and panels on digital libraries, authored over 40 peer-reviewed scientific articles, and has a forthcoming book *Affective Computing* to be published by MIT Press. She is a new associate editor of PAMI, with research interests in pattern modeling, continuous learning, and affective computing.

Alex Paul Pentland received his PhD from the Massachusetts Institute of Technology in 1982 and then began work at SRI International's Artificial Intelligence Center. He was appointed industrial lecturer in Stanford University's Computer Science Department in 1983, winning the Distinguished Lecturer award in 1986. In 1987 he returned to M.I.T. and is currently Head of the Perceptual Computing Section of the Media Laboratory, a group that includes over 50 researchers in computer vision, graphics, speech, music, and human–machine interaction. He has done research in artificial intelligence, machine vision, human vision, and computer graphics and has published more than 180 scientific articles in these areas. His most recent research focus is understanding human behavior in video, including face, expression, gesture, and intention recognition, as described in the April 1996 issue of *Scientific American*. He has won awards from the AAAI for his research into fractals; from the IEEE for his research into face recognition; and from Ars Electronica for his work in computer vision interfaces to virtual environments.