

## GUEST EDITORIAL PREFACE

# Content-Based Image Retrieval: Major Challenges for Biomedical Applications

*Thomas M. Deserno, RWTH Aachen University, Germany*

*L. Rodney Long, National Library of Medicine, NIH, USA*

*Hayit Greenspan, Tel Aviv University, Israel*

In most biomedical disciplines, digital image data is rapidly expanding in quantity and heterogeneity, and there is an increasing trend towards the formation of archives adequate to support diagnostics and preventive medicine. Exploration, exploitation, and consolidation of the immense image collections require tools to access structurally different data for research, diagnostics and teaching. Currently, image data is linked to textual descriptions, and data access is provided only via these textual additives. There are virtually no tools available to access medical images directly by their content or to cope with their structural differences. Therefore, visual-based (i.e. content-based) indexing and retrieval based on information contained in the pixel data of biomedical images is expected to have a great impact on biomedical image databases. However, existing systems for content-based image retrieval (CBIR) are not applicable to the biomedical imagery special needs, and novel methodologies are urgently needed.

This special issue grew from the workshop, *Content-Based Image Retrieval: Major Challenges for Medical Applications* at SPIE's International Symposium on Medical Imaging 2008 (Content-Based, 2008), which was convened

to assess status of CBIR within the biomedical clinical and research worlds, and to collect opinion from leading CBIR researchers about the most productive way forward. The workshop was structured around the concept of “gaps” (Deserno, Antani, & Long, 2008) between desired capabilities and use for medical CBIR, and what has actually been realized.

Panel participants in the workshop were asked to comment about a particular category of CBIR gap—content, features, performance, or usability—and also to provide informed opinion about critical gating factors affecting the use of CBIR in medicine (see (Deserno, et al., 2008) or the article by Long et al. in this issue for a discussion of these categories). Most of the articles in this issue are authored by panel participants and reflect this theme of “bridging gaps” in existing biomedical CBIR implementations.

Audience participants contributed a number of valuable comments, including one extended critique (see Acknowledgment) which we believe was particularly useful; we take the liberty to paraphrase it as “for CBIR to be valuable in medicine, know your data—images and related pathology and patient data—thoroughly; and incorporate the totality of this information into your

system". This same point is made independently and with elaboration in the article of Müller & Kalpathy-Cramer (in this issue).

In total, six articles that have passed the international review process were collected within this special issue. Brief introductions to each of the articles are given below.

Long et al. have provided an overview of the gap approach, illustrate its use in comparing four implemented CBIR systems, and argue for more attention in the CBIR research community to the areas of performance and usability, based on the relative inattention to these areas in the past, as measured by a survey of technical literature over the past seven years.

Horsthemke et al. contribute to understanding medical CBIR systems' semantic content and evaluation of performance with a focused investigation of mapping low-level mathematical shape-related features, including Fourier descriptors and a novel Radial Normal Index feature that they introduce, to the semantic features of spiculation, lobulation, and sphericity that are used by radiologists to describe shape of nodules in lung images. In their evaluation of the ability of the low-level features to predict the semantic interpretations of radiologists, the authors emphasize the significance of high variability of medical expert interpretation in establishing a reference "truth" standard for judging the success for algorithmic approaches in medicine.

Kim et al. describe two prototype applications for content-based image retrieval in a high-dimensional feature spaces. The first application is in 4D dynamic PET, and the second is dual-modal PET/CT images. These multi-dimensional imaging modalities are requiring new innovative techniques for content-based retrieval to fully utilize the massive amount of information in a meaningful and practical manner. The authors summarize research in bridging the feature gap for retrieval of multi-dimensional biomedical images.

Traina et al. focus on the performance gap in CBIR, including considerations that affect the user not only in perceived time of system response, but in effectiveness of the response in providing relevant information to satisfy the user's information goal. For example, they discuss techniques to reduce dimensionality of feature vectors not only with the goal of reducing

search time, but also to improve returned results as measured by precision vs. recall evaluations. In addition, they discuss and provide results to illustrate performance gains possible by database indexing of high-dimensional feature vectors, by analysis and optimization of query components, and by iterative relevance feedback from the user.

Greenspan addresses technical approaches to feature representation of image content and also the significant issue of representing not only global, but also local, region-of-interest content in medical images. She provides a detailed discussion of medical image retrieval with two approaches: (1) image landmark retrieval, where images are first segmented and key regions labeled; then the images are retrieved by labels; and (2) image-to-image retrieval, where an image is used as part or all of the query. She illustrates recent image content representation by the Gaussian Mixture Modeling-Kullback-Leibler (GMM-KL) method, as well as image "codebook" representation by image patches.

Müller & Kalpathy-Cramer specifically address the content gap in CBIR, contrasting the significance of purely visual image queries, which has been the main emphasis of biomedical CBIR researchers, versus queries that incorporate as much as possible of the complete patient context. They argue for the latter as a critical factor in advancing the use of CBIR in medicine. They contend that more than knowledge of imaging modality and viewing angle are required; rather, numerous case-specific contextual factors may be necessary to make images truly useful, and CBIR systems should address the need to incorporate these factors. Important factors that they provide for illustration in particular cases include patient age, medical goal of the image acquisition, and patient radiation treatment history.

In this issue we have provided expert responses for addressing specific problematic concerns in the field of medical CBIR, rather than a comprehensive survey. For such a survey, we refer the reader to the 2004 work of Müller, Michoux, Bandon, and Geissbuhler (2004). To those who have helped make this special issue possible by the contribution of their technical articles, and to all those researchers throughout the world who work to make and share progress in this highly challenging, sometimes frustrating,

and continually exciting enterprise of biomedical Content-based Image Retrieval, we express our deep gratitude.

## ACKNOWLEDGMENT

We thank Dr. Shih-Chung Ben Lo of the ISIS Center, Georgetown University, Washington, DC, for his thorough and helpful comments from the audience during the workshop.

## REFERENCES

Content-Based Image Retrieval: Major Challenges for Medical Applications. (2008, February). SPIE Medical Imaging. Presentations at: <http://archive.nlm.nih.gov/spiemi08/CBIRpanel.php>

Deserno, T. M., Antani, S., & Long, R. (2008, February). Ontology of gaps in content-based image retrieval. *Journal of Digital Imaging*. Online: <http://www.springerlink.com/content/?k=deserno>

Müller, H., Michoux, N., Bandon, D., & Geissbuhler, A. (2004). A review of content-based image retrieval systems in medical applications: Clinical benefits and future directions. *International Journal of Medical Informatics*. February, 73(1), 1-23.

*Thomas M. Deserno (born as Lehmann) received the diploma (masters degree) in electrical engineering, the PhD (summa cum laude) in computer science, and the habilitation in medical informatics from the RWTH Aachen University, Aachen, Germany, in 1992, 1998, and 2004, respectively. Since 2007, he is full professor for medical informatics. His research interests include medical image processing applied to quantitative measurements for computer-assisted diagnoses and content-based image retrieval from large medical databases. He currently serves on the international editorial board of Dentomaxillofacial Radiology as well as Methods of Information in Medicine, and he is co-editor of Europe of the International Journal of Healthcare Information Systems and Informatics.*

*L. Rodney Long is an electronics engineer for the Communications Engineering Branch at the National Library of Medicine, where he has worked since 1990. Prior to his current job, he worked for 14 years in industry as a software developer and as a systems engineer. His research interests are in telecommunications, image processing, and scientific/biomedical databases. He has an MA in applied mathematics from the University of Maryland.*

*Hayit Greenspan received the BSc and MSc degrees from the electrical engineering department of the Technion, Israel, in 1986 and 1989, respectively, and the PhD degree from the Electrical Engineering Dept. at CALTECH—California Institute of Technology, in 1994. Following the PhD she was a postdoc at the Computer Science Division at U.C. Berkeley. In 1997 she joined the Biomedical Engineering Dept. of the Faculty of Engineering, Tel-Aviv University, Israel, where she is currently a faculty member. Her research interests include medical image processing and analysis, content-based image and video search and retrieval, statistical image modeling and segmentation.*

## Editorial Review Board

Melanie S. Brodnik, Ohio State University, USA  
James A. Carter, PhD, Harvard Medical School, USA  
George Demiris, University of Missouri, USA  
Thomas M. Deserno, RWTH Aachen University, Germany  
Chengqi Guo, Mississippi State University, USA  
Thomas A. Horan, Claremont Graduate University, USA  
Chiehwen Ed Hsu, University of Maryland, USA  
Paul Hu, University of Utah, USA  
Patrick Hung, UOIT, Canada  
Richard Hull, University of Newcastle upon Tyne, United Kingdom  
Paul Keown, University of British Columbia, Canada  
Francis Lau, University of Victoria, Canada  
Cynthia LeRouge, St. Louis University, USA  
Huigang Liang, Florida Atlantic University, USA  
Shiyong Lu, Wayne State University, USA  
Phillip Olla, Madonna University, USA  
Liam O'Neil, UNT, USA  
Dave Parry, Auckland University of Technology, New Zealand  
Fay Cobb Payton, North Carolina State University, USA  
Ken Redekop, Erasmus Medical Center, The Netherlands  
J.P. Shim, Mississippi State University, USA  
Jim Warren, University of Auckland, New Zealand  
H. Joseph Wen, Southeast Missouri State University, USA  
Nilmini Wickramasinghe, Illinois Institute of Technology, USA  
Carla Wiggins, Idaho State University, USA  
Yufei Yuan, McMaster University, Canada  
Daniel Zeng, University of Arizona, USA  
Kai Zheng, The University of Michigan, USA

## Advisory Board

James G. Anderson, Purdue University, USA  
James Barlow, Imperial College of London, United Kingdom  
Marc D. Basson, John D. Dingell VA Medical Center; Wayne State University, USA  
William G. Chismar, University of Hawaii, USA  
Anand Desai, Ohio State University, USA  
Thomas W. Dillon, James Madison University, USA  
Guiseppe Forgionne, University of Maryland Baltimore County, USA  
Alan R. Hevner, University of South Florida, USA  
Nancy Lorenzi, Vanderbilt University Medical Center, USA  
Anthony F. Norcio, University of Maryland, USA  
Guy Paré, HEC Montréal, Canada  
Daniel Z. Sands, Harvard Medical School, USA  
Marion G. Sobol, Southern Methodist University, USA  
Brenda Zierler, University of Washington, USA

## Software/Book Review Editors

Kumar B. Keshavamurthy, Wayne State University, USA  
Joshia Tan, Washington University in St. Louis, USA

The *International Journal of Healthcare Information Systems and Informatics* is currently listed or indexed in: Burrelle's Media Directory; Cabell's Directory; CSA Illumina; GetCited; Google Scholar; The Index of Information Systems Journals; MediaFinder; Standard Periodical Directory; Ulrich's International Periodicals Directory.

# INTERNATIONAL JOURNAL OF HEALTHCARE INFORMATION SYSTEMS AND INFORMATICS

January-March 2009, Vol. 4, No. 1

## Table of Contents

SPECIAL ISSUE: CONTENT-BASED IMAGE RETRIEVAL

### GUEST EDITORIAL PREFACE

- i Content-Based Image Retrieval:  
Major Challenges for Biomedical Applications**  
*Thomas M. Deserno, RWTH Aachen University, Germany*  
*L. Rodney Long, National Library of Medicine, NIH, USA*  
*Hayit Greenspan, Tel Aviv University, Israel*

### RESEARCH ARTICLES

- 1 Content-Based Image Retrieval in Medicine:  
Retrospective Assessment, State of the Art, and Future Directions**  
*L. Rodney Long, National Library of Medicine, NIH, USA*  
*Sameer Antani, National Library of Medicine, NIH, USA*  
*Thomas M. Deserno, RWTH Aachen University, Germany*  
*George R. Thoma, National Library of Medicine, NIH, USA*
- 17 Evaluation Challenges for Bridging Semantic Gap:  
Shape Disagreements on Pulmonary Nodules in the Lung Image Database  
Consortium**  
*William H. Horsthemke, DePaul University, USA*  
*Daniela S. Raicu, DePaul University, USA*  
*Jacob D. Furst, DePaul University, USA*
- 34 Bridging the Feature Gaps for Retrieval of Multi-Dimensional Images**  
*Jinman Kim, University of Sydney, Australia, and Polytechnic University, Hong Kong*  
*Weidong Cai, University of Sydney, Australia, and Polytechnic University, Hong Kong*  
*Dagan Feng, University of Sydney, Australia, and Polytechnic University, Hong Kong*
- 47 How to Cope with the Performance Gap in Content-Based Image Retrieval Systems**  
*Agma J. M. Traina, University of São Paulo at São Carlos, Brazil*  
*Caetano Traina Jr., University of São Paulo at São Carlos, Brazil*  
*Cristina D. A. Ciferri, University of São Paulo at São Carlos, Brazil*  
*Marcela X. Ribeiro, University of São Paulo at São Carlos, Brazil*  
*Paulo M. Azevedo-Marques, University of São Paulo at Ribeirão Preto, Brazil*
- 68 Revisiting the Feature and Content Gap for Landmark-Based and Image-to-Image  
Retrieval in Medical CBIR**  
*Hayit Greenspan, Tel-Aviv University, Israel*
- 88 Putting the Content Into Context: Features and Gaps in Image Retrieval**  
*Henning Müller, University and Hospitals of Geneva and University of Applied Sciences,  
Switzerland*  
*Jayashree Kalpathy-Cramer, Oregon Health and Science University, USA*

# International Journal of Healthcare Information Systems and Informatics

## Guidelines for Manuscript Submissions

### Mission

The **International Journal of Healthcare Information Systems & Informatics (IJHISI)** is a quarterly, peer-reviewed scholarly journal, which promotes (advance) research in the application of information systems and informatics to the healthcare arena. The mission of this journal is to act as a primer source for both academic researchers and practitioners to advance the knowledge and practice in the interdisciplinary areas of healthcare information technology and management research and to address the innovative concepts and critical issues in the emerging field of health information systems and informatics.

### Coverage

As the purpose of this journal is to bring together experts in public health, health information systems, and medical informatics to establish a new paradigm for developing health information systems & informatics, topics covered will span a wide scope of applications. Issues surrounding the organization, classification, evaluation and adoption of advanced information and communication technology in: Healthcare; IT security and privacy issues; Clinical decision support design, development and implementation; Management of emerging healthcare technologies; Role of informatics specialists; Telemedicine; E-Health; M-Health; and much more! Visit [www.igi-global.com/ijhisi](http://www.igi-global.com/ijhisi).

While it is anticipated that the majority of contributing authors will be from the IS groups interested in the applications of IS in healthcare, the journal will also be open to contributions from the specialty medical and clinical fields, including clinical decision support systems, nursing expert systems, pharmaceutical and home healthcare informatics, IT applications in physical therapeutic treatments, rehabilitative technologies and telemedicine. By taking a broad scope with a multidisciplinary approach to understanding HISI and its implications on governmental policies, public and private sector industries, and the global economies, this journal promises to cumulate and integrate the theories, methodologies and best practices in IJHISI.

### Originality

Only original manuscripts submitted to IJHISI will be considered for publication. The cover letter should include a statement that the findings reported in the manuscript have not been previously published and that the manuscript is not being simultaneously submitted elsewhere. If the manuscript (or any version of it) has appeared in conference proceedings, details of such publication must be made known to the Editor at the time of submission so that the suitability of the manuscript for IJHISI can be assessed.

### Style

Prepare manuscripts according to the Publication Manual of the American Psychological Association, paying particular attention to references and citations and following "Guidelines to Reduce Bias in Language" (APA Manual, pp. 46-60). Include: (a) authors' last names and year of publication in text citations and (b) complete references that correspond to the citations. In references, spell out journal titles and do not use abbreviations. See the following examples:

#### Example 1: Journal Article.

Trippi, R., & Turban, E. (1989). The impact of parallel and neural computing on managerial decision making. *Journal of Management Information Systems*, 6, 85-98.

#### Example 2: Entire Book

Goldberg, D. E. (1989). *Genetic algorithms in search, optimization, and machine learning*. Reading, MA: Addison-Wesley.

#### Example 3: Book Chapter

Yamaguchi, T., Goto, K., & Takagi, T. (1994). Fuzzy neural net system. In A. L. Ralescu (Ed.), *Applied research in fuzzy technology* (pp. 295-369). Boston: Kluwer.

#### Example 4: Proceedings Paper

Ritter, H., & Schulten, K. (1988). Kohonen's self-organizing maps: Exploring their computational capabilities. In *Proceedings of the IEEE Second International Conference on Neural Networks, I* (pp. 109-116).

**Footnotes/Endnotes** are not used in IJHISI.

**Mathematical Symbols:** Submit a list of symbols used in mathematical formulas, including a description of each symbol, and clearly mark in the text and margins, any formulas that require boldface print.

**Tables:** Keep each table to a reasonable size (very large tables packed with data simply overwhelm the reader). Give each table and each column an explanatory heading. Clearly indicate units of measurement. Do not reproduce same data in both tables and figures (illustrations). Prepare table footnotes according to APA guidelines. Consecutively number tables. Make tables completely intelligible without reference to text. If table is reprinted or adapted from another source, include credit line.

**Figures and Figure Captions:** Make figures for reproduction about twice the final size required so that the reduced figures will fit on the journal page; make all type equal in size and no higher than .25 in. Figures should be: (a) high-quality, professionally drawn, high-resolution, computer-generated illustrations or (b) glossy, high-contrast, black-and-white photographs. (Authors of articles requiring use of four-color figures should contact the publisher.) Do not trim, mount, clip, staple, or write on back of figures. Instead, write article title, figure number, and TOP (of figure) on label and apply label to back of each figure. Consecutively number figures. Securely package figures in a protective envelope. Attach photocopies of all figures to manuscript. Consecutively number captions with Arabic numerals corresponding to the figure numbers. Make captions completely intelligible without reference to text. If figure is reprinted or adapted from another source, include credit line.

### **Review Process**

The appropriate editor will examine each manuscript to make a preliminary judgment of its relevance and quality. If deemed of potential value, the manuscript will be sent to at least 2 reviewers for their comments. The amount of time taken to reach a decision will be kept to a minimum and the total time will be tracked to ensure that there is no unnecessary delay in the evaluation procedure.

### **Copyright**

Authors are responsible for all statements made in their work and for obtaining permission to reprint or adapt a copyrighted table or figure or to quote at length from a copyrighted work. Authors should write to original author(s) and original publisher to see if permission is required and to request nonexclusive world rights in all languages to use the material in the current article and in future editions. Include copies of all permissions and credit lines with the manuscript.

### **Submission**

Interested authors are asked to submit their manuscript as an attachment to an e-mail. The manuscript must be in Word or RTF (Rich Text Format) to e-mail: an9188@wayne.edu. The title page should contain the paper's title, author(s)' name and affiliation, mailing address, phone number, fax number, and email address of each author. The authors' names and affiliations should NOT appear in the manuscript, except on the title page. The second page, should contain the article title and abstract (no more than 150 words). The abstract should include the purpose of the study, basic procedures, most important findings, and main conclusions, with emphasis on new aspects of the study.

### **Length**

The length of the submitted manuscript must not exceed 32 double-spaced, typed pages, including table and figure pages. The following guidelines must be met in terms of length: Only use front side of paper Each component begins on a new page Abbreviated article title and page number is typed in the upper right-hand corner of each page Order of title page, abstract and key words, text, references, appendices, tables, and figure captions

### **Correspondence**

The acknowledgment letter regarding the receipt of the manuscript will be promptly sent. The review process will take approximately 8-16 weeks, and the author will be notified concerning the possibility of publication of the manuscript as soon as the review process is completed. All correspondence will be directed to the first author of multi-authored manuscripts. It is the responsibility of the first author to communicate with the other author(s). Authors of accepted manuscript will be asked to provide a final copy of their manuscript in either Word or RTF text format stored on a 3 1/2" disk, zip disk or CD-ROM, accompanied by a hardcopy of the manuscript and the original signed copy of the Warranty and Copyright Agreement. The accepted manuscript will be edited by the Journal copyeditor for format and style.

### **All submissions and inquiries should be directed to the attention of:**

Joseph Tan, PhD  
Wayne State University  
Information Systems & Manufacturing  
School of Business  
Editor-in-Chief, IJHISI  
E-Mail: an9188@wayne.edu