PROCEEDINGS OF

Electronic Imaging

Science and Technology

Document Recognition and Retrieval XIV

Xiaofan Lin
Berrin A. Yanikoglu
Chairs/Editors

30 January–1 February 2007
San Jose, California, USA

Sponsored by
IS&T—The Society for Imaging Science and Technology
SPIE—The International Society for Optical Engineering

SPIE Vol. 6500

The papers included in this volume were part of the technical conference cited on the cover and title page. Papers were selected and subject to review by the editors and conference program committee. Some conference presentations may not be available for publication. The papers published in these proceedings reflect the work and thoughts of the authors and are published herein as submitted. The publishers are not responsible for the validity of the information or for any outcomes resulting from reliance thereon.

Please use the following format to cite material from this book:
Author(s), "Title of Paper," in *Document Recognition and Retrieval XIV*, edited by Xiaofan Lin, Berrin A. Yanikoglu, Proceedings of SPIE-IS&T Electronic Imaging, SPIE Vol. 6500, Article CID Number (2007).

ISSN 0277-786X
ISBN 9780819466136

Copublished by
SPIE—The International Society for Optical Engineering
P.O. Box 10, Bellingham, Washington 98227-0010 USA
Telephone 1 360/676-3290 (Pacific Time) · Fax 1 360/647-1445
http://www.spie.org
and
IS&T—The Society for Imaging Science and Technology
7003 Kilworth Lane, Springfield, Virginia, 22151 USA
Telephone 1 703/642-9090 (Eastern Time) · Fax 1 703/642-9094
http://www.imaging.org

Printed in the United States of America.

Contents

Pagination: Proceedings of SPIE follow an e-First publication model, with papers published first online and then in print and on CD-ROM. Papers are published as they are submitted and meet publication criteria. A unique, consistent, permanent citation identifier (CID) number is assigned to each article at the time of the first publication. Utilization of CIDs allows articles to be fully citable as soon they are published online, and connects the same identifier to all online, print, and electronic versions of the publication.

SPIE uses a six-digit CID article numbering system in which:
• The first four digits correspond to the SPIE volume number.
• The last two digits indicate publication order within the volume using a Base 36 numbering system employing both numerals and letters. The CID number appears on each page of the manuscript. The complete citation is used on the first page, and an abbreviated version on subsequent pages.

Conference Committee

Symposium Chairs

Michael A. Kriss, Consultant (USA)
Robert A. Sprague, Consultant (USA)

Conference Chairs

Xiaofan Lin, Riya Inc. (USA)
Berrin A. Yanikoglu, Sabanci University (Turkey)

Program Committee

Jan P. Allebach, Purdue University (USA)
Tim L. Andersen, Boise State University (USA)
Apostolos Antonacopoulos, University of Salford (United Kingdom)
Elisa H. Barney Smith, Boise State University (USA)
Kathrin Berkner, Ricoh Innovations, Inc. (USA)
Hui Chao, Hewlett-Packard Company (USA)
Brian D. Davison, Lehigh University (USA)
Xiaoqing Ding, Tsinghua University (China)
David S. Doermann, University of Maryland, College Park (USA)
Steven J. Harrington, Xerox Corporation (USA)
Jianying Hu, IBM Thomas J. Watson Research Center (USA)
Matthew F. Hurst, Intelliseek, Inc. (USA)
Hisashi Ikeda, Hitachi, Ltd. (Japan)
Tapas Kanungo, IBM Almaden Research Center (USA)
Daniel P. Lopresti, Lehigh University (USA)
Thomas A. Nartker, University of Nevada, Las Vegas (USA)
Sargur N. Srihari, SUNY, University at Buffalo (USA)
Kazem Taghva, University of Nevada, Las Vegas (USA)
George R. Thoma, National Library of Medicine (USA)

Introduction

Information manifests itself in various media and many forms. The goal of the Document Recognition and Retrieval Conference is to provide a forum to bring together researchers working on all stages and various aspects of converting paper documents into searchable and repurposable media, spanning a wide-area including document recognition (converting document images into electronic form), keyword spotting (retrieving document images containing keywords), information extraction and retrieval from documents, and document repurposing.

The papers in this volume represent the state-of-the-art in diverse areas such as handwriting recognition, optical character recognition, document image processing, information retrieval, and classification. In addition, this year we have six papers in the Digital Publishing Special Sessions.

We are fortunate to have two invited talks by eminent veterans in OCR industry. Dr. István Marosi of Nuance Communications will talk on "Industrial OCR approaches: architecture, algorithms, and adaptation techniques" and Mr. Jim Fruchterman of Benetech will talk on "Document recognition serving people with disabilities."

New for this year, the Best Student Paper will be selected among papers whose lead authors are full-time students. We gratefully acknowledge Nuance Communications, the worldwide leading provider of imaging and speech technologies and solutions, for generously sponsoring this award.

We thank the program committee members and the SPIE conference organizers for their help in organizing the conference and for their help in the review process. We especially thank the participating researchers for their contributions to this cross-disciplinary conference.

As always, we want to hear from readers of this volume, and we encourage you to approach the conference co-chairs or any member of our distinguished program committee with suggestions for future improvements.

We hope everyone finds this year's conference simulating and rewarding, with its unique opportunities for interaction.

Xiaofan Lin
Berrin A. Yanikoglu