

# ***Document Recognition and Retrieval XVI***

**Kathrin Berkner  
Laurence Likforman-Sulem**  
*Editors*

**20–22 January 2009  
San Jose, California, USA**

*Sponsored and Published by*  
IS&T—The Society for Imaging Science and Technology  
SPIE

*Cosponsored by*  
Ricoh Innovations, Inc. (USA)

**Volume 7247**

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 XVI*, edited by Kathrin Berkner, Laurence Likforman-Sulem, Proceedings of SPIE-IS&T Electronic Imaging, SPIE Vol. 7247, Article CID Number (2009).

ISSN 0277-786X  
ISBN 9780819474971

Copublished by

**SPIE**

P.O. Box 10, Bellingham, Washington 98227-0010 USA  
Telephone +1 360 676 3290 (Pacific Time) · Fax +1 360 647 1445  
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  
imaging.org

Copyright © 2009, Society of Photo-Optical Instrumentation Engineers and The Society for Imaging Science and Technology.

Copying of material in this book for internal or personal use, or for the internal or personal use of specific clients, beyond the fair use provisions granted by the U.S. Copyright Law is authorized by the publishers subject to payment of copying fees. The Transactional Reporting Service base fee for this volume is \$18.00 per article (or portion thereof), which should be paid directly to the Copyright Clearance Center (CCC), 222 Rosewood Drive, Danvers, MA 01923. Payment may also be made electronically through CCC Online at [copyright.com](http://copyright.com). Other copying for republication, resale, advertising or promotion, or any form of systematic or multiple reproduction of any material in this book is prohibited except with permission in writing from the publisher. The CCC fee code is 0277-786X/09/\$18.00.

Printed in the United States of America.

---

**Paper Numbering:** 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. These two-number sets start with 00, 01, 02, 03, 04, 05, 06, 07, 08, 09, 0A, 0B ... 0Z, followed by 10-1Z, 20-2Z, etc.

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. Numbers in the index correspond to the last two digits of the six-digit CID number.

# Contents

vii	<i>Conference Committee</i>
ix	<i>Introduction</i>

---

## SESSION 1 INVITED PRESENTATION

---

- 7247 02 **Pseudo-color enhanced x-ray fluorescence imaging of the Archimedes Palimpsest (Invited Paper)** [7247-01]  
U. Bergmann, SLAC National Accelerator Lab. (United States); K. T. Knox, Boeing LTS (United States)

---

## SESSION 2 SEGMENTATION

---

- 7247 03 **Text-image alignment for historical handwritten documents** [7247-02]  
S. Zinger, Eindhoven Univ. of Technology (Netherlands); J. Nerbonne, L. Schomaker, Univ. of Groningen (Netherlands)
- 7247 04 **Document boundary determination using structural and lexical analysis** [7247-04]  
K. Taghva, M.-A. Cartright, Univ. of Nevada, Las Vegas (United States)
- 7247 05 **Segmentation of continuous document flow by a modified backward-forward algorithm** [7247-05]  
Th. Meilender, A. Belaïd, Univ. Nancy 2 - LORIA (France)

---

## SESSION 3 RETRIEVAL AND TEXT CATEGORIZATION

---

- 7247 06 **Retrieval of historical documents by word spotting** [7247-06]  
N. Doulgéri, E. Kavallieratou, Univ. of the Aegean (Greece)
- 7247 07 **Enriching a document collection by integrating information extraction and PDF annotation** [7247-07]  
B. Powley, R. Dale, I. Anisimoff, Macquarie Univ. (Australia)
- 7247 08 **Locating and parsing bibliographical references in HTML medical articles** [7247-08]  
J. Zou, D. Le, G. R. Thoma, National Library of Medicine (United States)
- 7247 09 **On-line handwritten text categorization** [7247-09]  
S. Peña Saldarriaga, LINA, UMR CNRS 6241, Univ. de Nantes (France); C. Viard-Gaudin, IRCCyn, UMR CNRS 6597, Univ. de Nantes (France); E. Morin, LINA, UMR CNRS 6241, Univ. de Nantes (France)

---

**SESSION 4 RECOGNITION I**

---

- 7247 0A **Improvement of Arabic handwriting recognition systems: combination and/or reject?** [7247-10]  
H. El Abed, V. Märgner, Technical Univ. Braunschweig (Germany)
- 7247 0B **A robust model for on-line handwritten Japanese text recognition** [7247-11]  
B. Zhu, Tokyo Univ. of Agriculture and Technology (Japan); X.-D. Zhou, C.-L. Liu, Institute of Automation (China); M. Nakagawa, Tokyo Univ. of Agriculture and Technology (Japan)
- 7247 0C **Online computation of similarity between handwritten characters** [7247-12]  
O. Golubitsky, S. M. Watt, Univ. of Western Ontario (Canada)

---

**SESSION 5 INVITED PRESENTATION**

---

- 7247 0D **Advanced topics in character recognition and document analysis: research works in intelligent image and document research lab, Tsinghua University (Invited Paper)** [7247-13]  
X. Ding, Tsinghua Univ. (China)

---

**SESSION 6 WRITER OR SCRIPT IDENTIFICATION**

---

- 7247 0E **Comparison of statistical models for writer verification** [7247-14]  
S. Srihari, G. R. Ball, Univ. at Buffalo (United States)
- 7247 0F **Online writer identification using alphabetic information clustering** [7247-16]  
G. X. Tan, Nanyang Technological Univ. (Singapore) and IRCCyN, UMR CNRS 6597, Univ. de Nantes (France); C. Viard-Gaudin, IRCCyN, UMR CNRS 6597, Univ. de Nantes (France); A. C. Kot, Nanyang Technological Univ. (Singapore)

---

**SESSION 7 RECOGNITION II**

---

- 7247 0G **Using synthetic data safely in classification** [7247-17]  
J. Nonnemaker, H. S. Baird, Lehigh Univ. (United States)
- 7247 0H **Combination of dynamic Bayesian network classifiers for the recognition of degraded characters** [7247-18]  
L. Likforman-Sulem, M. Sigelle, TELECOM ParisTech/TSI and CNRS LTCI (France)
- 7247 0I **Character recognition in the presence of occluding clutter** [7247-19]  
K. T. Fosseide, L. Aurdal, Lumex AS (Norway)
- 7247 0J **Multi-font printed Mongolian document recognition system** [7247-20]  
L. Peng, C. Liu, X. Ding, H. Wang, J. Jin, Tsinghua Univ. (China)

---

**SESSION 8 SEGMENTATION AND RESTORATION**

---

- 7247 OK **Resolution independent skew and orientation detection for document images** [7247-21]  
J. van Beusekom, Technical Univ. of Kaiserslautern (Germany); F. Shafait, DFKI GmbH (Germany); T. M. Breuel, DFKI GmbH (Germany) and Technical Univ. of Kaiserslautern (Germany)
- 7247 OL **Text line extraction in free style document** [7247-22]  
X. Shen, C. Liu, X. Ding, Tsinghua Univ. (China); Y. Zou, Nokia Research Ctr. (China)
- 7247 OM **Simultaneous detection of vertical and horizontal text lines based on perceptual organization** [7247-23]  
C. Faure, CNRS-LTCl, TELECOM ParisTech (France); N. Vincent, CRIP5-Univ. Paris Descartes (France)
- 7247 ON **Efficient shape-LUT classification for document image restoration** [7247-24]  
T. Obafemi-Ajayi, G. Agam, O. Frieder, Illinois Institute of Technology (United States)

---

**SESSION 9 IMAGE PROCESSING**

---

- 7247 OO **Camera-based document image mosaicing using LLAH** [7247-25]  
T. Nakai, K. Kise, M. Iwamura, Osaka Prefecture Univ. (Japan)
- 7247 OP **Mark detection from scanned ballots** [7247-26]  
E. H. Barney-Smith, Boise State Univ. (United States); G. Nagy, Rensselaer Polytechnic Institute (United States); D. Lopresti, Lehigh Univ. (United States)

---

**INTERACTIVE PAPER SESSION**

---

- 7247 OQ **Improving semi-text-independent method of writer verification using difference vector** [7247-27]  
X. Li, X. Ding, Tsinghua Univ. (China)
- 7247 OR **Restoring warped document image through segmentation and full page interpolation** [7247-28]  
Y. Zhang, C. Liu, X. Ding, Tsinghua Univ. (China); K. Wang, Nokia Research Ctr. (China)
- 7247 OS **Identification of forgeries in handwritten petitions for ballot propositions** [7247-29]  
S. Srihari, V. Ramakrishnan, M. Malgireddy, G. R. Ball, Univ. at Buffalo (United States)
- 7247 OT **Simultaneous segmentation and recognition of Arabic printed text using linguistic concepts of vocabulary** [7247-30]  
M. Ben Halima, A. M. Alimi, The High School of National Engineering of Sfax (Tunisia)
- 7247 OU **Comparison of Niblack inspired binarization methods for ancient documents** [7247-31]  
K. Khurshid, I. Siddiqi, Lab. CRIP5-SIP, Univ. Paris Descartes (France); C. Faure, UMR CNRS 5141, GET ENST (France); N. Vincent, Univ. Paris Descartes (France)

- 7247 OV **Figure content analysis for improved biomedical article retrieval** [7247-34]  
D. You, The State Univ. of New York, Buffalo (United States); E. Apostolova, DePaul Univ. (United States); S. Antani, D. Demner-Fushman, G. R. Thoma, National Library of Medicine (United States)
- 7247 OW **A semi-supervised learning method to classify grant support zone in web-based medical articles** [7247-35]  
X. Zhang, J. Zou, D. X. Le, G. Thoma, National Library of Medicine (United States)
- 7247 OX **Layout-free dewarping of planar document images** [7247-36]  
M. Iwamura, R. Niwa, A. Horimatsu, K. Kise, Osaka Prefecture Univ. (Japan); S. Uchida, Kyushu Univ. (Japan); S. Omachi, Tohoku Univ. (Japan)
- 7247 OY **Watermarking ancient documents based on wavelet packets** [7247-38]  
M. N. Maatouk, Faculty of Sciences of Monastir (Tunisia); O. Jedidi, National Engineering School of Monastir (Tunisia); N. Essoukri Ben Amara, National Engineering School of Sousse (Tunisia)
- 7247 OZ **Script identification of handwritten word images** [7247-15]  
A. Bhardwaj, H. Cao, V. Govindaraju, Univ. at Buffalo (United States)

*Author Index*

# Conference Committee

## *Symposium Chairs*

**Nitin Sampat**, Rochester Institute of Technology (United States)  
**Jan P. Allebach**, Purdue University (United States)

## *Conference Chairs*

**Kathrin Berkner**, Ricoh Innovations, Inc. (United States)  
**Laurence Likforman-Sulem**, TELECOM ParisTech (France)

## *Program Committee*

**Gady Agam**, Illinois Institute of Technology (United States)  
**Tim L. Andersen**, Boise State University (United States)  
**Apostolos Antonacopoulos**, University of Salford (United Kingdom)  
**Elisa H. Barney-Smith**, Boise State University (United States)  
**Xiaoqing Ding**, Tsinghua University (China)  
**David S. Doermann**, University of Maryland, College Park (United States)  
**Jianying Hu**, IBM Thomas J. Watson Research Center (United States)  
**Matthew F. Hurst**, Intelliseek, Inc. (United States)  
**Tapas Kanungo**, Yahoo! Inc. (United States)  
**Daniel P. Lopresti**, Lehigh University (United States)  
**Lambert R. B. Schomaker**, University of Groningen (Netherlands)  
**Xiaofan Lin**, Riya, Inc. (United States)  
**Hiroshi Sako**, Hitachi, Ltd. (Japan)  
**Sargur N. Srihari**, University at Buffalo (United States)  
**Venkata Subramaniam**, IBM India Research Laboratory (India)  
**Kazem Taghva**, University of Nevada, Las Vegas (United States)  
**George R. Thoma**, National Library of Medicine (United States)  
**Alessandro Vinciarelli**, IDIAP Research Institute (Switzerland)  
**Berrin Yanikoglu**, Sabanci University (Turkey)



## Introduction

This volume contains papers presented at the 16th Document Recognition and Retrieval Conference held in January 2009, in San Jose, CA. The conference is part of the Electronic Imaging Symposium, which brings together researchers from various backgrounds related to electronic imaging for an exciting research event.

In that light we are fortunate to include invited talks on two very diverse topics into our conference. Prof. Xiaoqing Ding of Tsinghua University (China) will give an overview for the field of recognition of modern handwritten documents in Asian languages, whereas Uwe Bergmann and Keith Knox will speak about their work on deciphering a rather ancient document, the Archimedes Palimpsest.

Framed by these two presentations, this volume contains state-of-the-art research covering areas such as document recognition in presence of degradations, handwriting recognition, document image processing, and writer identification.

This year we continue the tradition of giving an award for the Best Student Paper to a paper whose lead author is a full-time student. We gratefully acknowledge Ricoh Innovations for generously sponsoring this award.

We thank the program committee of DRR and the SPIE conference organizers for their help in organizing the conference and facilitating the review. Many thanks go also to the external reviewers for assisting in the review process. At last, we thank all the participating authors of the papers for their contributions to this conference.

As always, we welcome any feedback from readers of this volume and we encourage you to approach any member of the program committee with suggestions for further improvements. The landscape of document processing is constantly changing with a lot of research questions to be discovered in the future. We encourage you to actively search for those questions and use this conference as an opportunity for discussions on your topic of interest with colleagues in a unique interdisciplinary setting.

**Kathrin Berkner**  
**Laurence Likforman-Sulem**

