

Communications in Computer and Information Science

1020

Commenced Publication in 2007

Founding and Former Series Editors:

Phoebe Chen, Alfredo Cuzzocrea, Xiaoyong Du, Orhun Kara, Ting Liu,
Krishna M. Sivalingam, Dominik Ślęzak, Takashi Washio, and Xiaokang Yang

Editorial Board Members

Simone Diniz Junqueira Barbosa

*Pontifical Catholic University of Rio de Janeiro (PUC-Rio),
Rio de Janeiro, Brazil*

Joaquim Filipe

Polytechnic Institute of Setúbal, Setúbal, Portugal

Ashish Ghosh

Indian Statistical Institute, Kolkata, India

Igor Kotenko

*St. Petersburg Institute for Informatics and Automation of the Russian
Academy of Sciences, St. Petersburg, Russia*

Junsong Yuan

University at Buffalo, The State University of New York, Buffalo, NY, USA

Lizhu Zhou

Tsinghua University, Beijing, China

More information about this series at <http://www.springer.com/series/7899>

Suresh Sundaram · Gaurav Harit (Eds.)

Document Analysis and Recognition

4th Workshop, DAR 2018

Held in Conjunction with ICVGIP 2018

Hyderabad, India, December 18, 2018

Revised Selected Papers

Editors

Suresh Sundaram
Electronics and Electrical Engineering
Indian Institute of Technology Guwahati
Guwahati, India

Gaurav Harit
Computer Science and Engineering
Indian Institute of Technology Jodhpur
Karwar, Rajasthan, India

ISSN 1865-0929

ISSN 1865-0937 (electronic)

Communications in Computer and Information Science

ISBN 978-981-13-9360-0

ISBN 978-981-13-9361-7 (eBook)

<https://doi.org/10.1007/978-981-13-9361-7>

© Springer Nature Singapore Pte Ltd. 2019

This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

The publisher, the authors and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, expressed or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

This Springer imprint is published by the registered company Springer Nature Singapore Pte Ltd.
The registered company address is: 152 Beach Road, #21-01/04 Gateway East, Singapore 189721, Singapore

Preface

The 4th Workshop on Document Analysis and Recognition (DAR 2018) was held at Hyderabad, Telangana, India, on December 18, 2018. This edition was organized by the International Institute of Information Technology, Hyderabad (IIIT-H) as part of the 11th Indian Conference on Vision, Graphics, and Image Processing (ICVGIP 2018). The DAR workshop series brings together researchers working in the area of document analysis, with the focus of exchanging ideas in order to foster further research in the field. The present proceedings contain the papers that were accepted and presented at the workshop.

Broadly speaking, the papers presented in this volume describe developments in different facets of document analysis, such as:

- (a) Document layout analysis and understanding
- (b) Handwriting recognition and symbol spotting
- (c) Character and word segmentation
- (d) Handwriting analysis
- (e) Datasets and performance evaluation

From a total of 22 papers submitted to the workshop, a total of 14 were accepted and presented. The papers were selected based on three reviews obtained for each of the submissions. Members of the Program Committee were involved in the review of the papers, for which a single-blinded procedure was adopted. The papers in the proceedings are the revised versions that were submitted after the incorporation of the reviewer comments. Apart from paper presentations, the workshop also hosted a keynote talk by Prof. A. G. Ramakrishnan, Department of Electrical Engineering, Indian Institute of Science, Bangalore.

June 2019

Suresh Sundaram
Gaurav Harit

Organization

Advisory Committee

Bidyut Baran Chaudhuri	ISI Kolkata, India
Santanu Chaudhury	CSIR-CEERI and IIT Delhi, India
Umapada Pal	ISI Kolkata, India
Angarai Ganesan	IISc Bangalore, India
Ramakrishnan	

Program Co-chairs

Suresh Sundaram	IIT Guwahati, India
Gaurav Harit	IIT Jodhpur, India

Program Committee

Ujjwal Bhattacharya	ISI Kolkata, India
Utpal Garain	ISI Kolkata, India
Prithwijit Guha	IIT Guwahati, India
D. S. Guru	University of Mysore, India
Thotreingam Kasar	Kaaya Tech, India
Gurpreet Singh Lehal	Panjab University, India
Anand Mishra	IISc Bangalore, India
Atul Negi	University of Hyderabad, India
Swapan Parui	ISI Kolkata, India
Shivakumara Palaiahnakote	University of Malaya, Malaysia
Sitaram Ramachandrula	DXC Technology, India
Partha Pratim Roy	IIT Roorkee, India
Rajendra Kumar Sharma	Thapar Institute of Engineering and Technology, India
Kiran Hiremath	IIT Jodhpur, India
Yashaswi Verma	IIT Jodhpur, India
Abhishek Sharma	IIIT Naya Raipur, India
A. V. Narasimhadhan	NIT Karnataka, India
Ayesha Choudhury	Jawaharlal Nehru University, India

Contents

Document Layout Analysis and Understanding

MultiDIAS: A Hierarchical Multi-layered Document Image Annotation System	3
<i>Arnab Poddar, Rohan Mukherjee, Jayanta Mukhopadhyay, and Prabir Kumar Biswas</i>	
Attributed Paths for Layout-Based Document Retrieval	15
<i>Divya Sharma, Gaurav Harit, and Chiranjoy Chattopadhyay</i>	
Textual Content Retrieval from Filled-in Form Images.	27
<i>Soulis Ghosh, Rajdeep Bhattacharya, Sandipan Majhi, Showmik Bhowmik, Samir Malakar, and Ram Sarkar</i>	

Handwriting Recognition and Symbol Spotting

A Study on the Effect of CNN-Based Transfer Learning on Handwritten Indic and Mixed Numeral Recognition.	41
<i>Rahul Pramanik, Prabhat Dansena, and Soumen Bag</i>	
Symbol Spotting in Offline Handwritten Mathematical Expressions.	52
<i>Ridhi Aggarwal, Gaurav Harit, and Anil Kumar Tiwari</i>	
Online Handwritten Bangla Character Recognition Using Frechet Distance and Distance Based Features	65
<i>Shibaprasad Sen, Jewel Chakraborty, Snehanjan Chatterjee, Rohit Mitra, Ram Sarkar, and Kaushik Roy</i>	
An Efficient Multi Lingual Optical Character Recognition System for Indian Languages Through Use of Bharati Script.	74
<i>Chandra Sekhar Vorugunti, Srinivasa Chakravarthy, and Viswanath Pulabaigari</i>	

Character and Word Segmentation

Telugu Word Segmentation Using Fringe Maps	87
<i>Koteswara Rao Devarapalli and Atul Negi</i>	
An Efficient Character Segmentation Algorithm for Connected Handwritten Documents.	97
<i>Vishal Rajput, N. Jayanthi, and S. Indu</i>	

Handwriting Analysis

A Deep Learning Architecture Based Dimensionality Reduction and Online Signature Verification	109
--	-----

Chandra Sekhar Vorugunti and Viswanath Pulabaigari

Word-Wise Handwriting Based Gender Identification Using Multi-Gabor Response Fusion	119
--	-----

*Maryam Asadzadeh Kaljahi, P. V. Vidya Varshini,
Palaiahnakote Shivakumara, Umapada Pal, Tong Lu, and D. S. Guru*

A Secure and Light Weight User Authentication System Based on Online Signature Verification for Resource Constrained Mobile Networks	133
---	-----

Chandra Sekhar Vorugunti, D. S. Guru, and Viswanath Pulabaigari

Datasets and Performance Evaluation

Benchmark Datasets for Offline Handwritten Gurmukhi Script Recognition.	143
--	-----

*Munish Kumar, R. K. Sharma, M. K. Jindal, Simpel Rani Jindal,
and Harjeet Singh*

Benchmark Dataset: Offline Handwritten Gurmukhi City Names for Postal Automation	152
---	-----

Harmandeep Kaur and Munish Kumar

Author Index	161
------------------------	-----