ICDAR 2013

12th International Conference on Document Analysis and Recognition

25–28 August 2013
Washington, DC

Published by:
CPS
Conference Publishing Services

Conference Information | Papers by Session | Papers by Author
Getting Started | Trademarks | Search

Proceedings

12th International Conference on Document Analysis and Recognition

ICDAR 2013

25–28 August 2013
Washington, DC

2013 12th International Conference on Document Analysis and Recognition

ICDAR 2013

Table of Contents

Welcome from the Executive Chairs xxiv

Welcome from the Program Chairs xxvi

Organizing Committee xxvii

Program Committee xxviii

Sponsors xxix

Keynotes xxx

Oral Session 1: Character Recognition 1

Analyzing the Distribution of a Large-Scale Character Pattern Set Using Relative Neighborhood Graph 3

Masanori Goto, Ryosuke Ishida, Yaokai Feng, and Seiichi Uchida

Locally Smoothed Modified Quadratic Discriminant Function 8

Xu-Yao Zhang and Cheng-Lin Liu

IBM_UB_1: A Dual Mode Unconstrained English Handwriting Dataset 13

Arti Shivram, Chetan Ramaiah, Srirangaraj Setlur, and Venu Govindaraju

Character Recognition Using Conditional Random Field Based Recognition Engine 18

Anupama Ray, Ankit Chandawala, and Santanu Chaudhury

Oral Session 2: Applications 1

The Wordometer—Estimating the Number of Words Read Using Document Image Retrieval and Mobile Eye Tracking 25

Kai Kunze, Hitoshi Kawaichi, Kazuyo Yoshimura, and Koichi Kise

Wearable Reading Assist System: Augmented Reality Document Combining Document Retrieval and Eye Tracking 30

Takumi Toyama, Andreas Dengel, Wakana Suzuki, and Koichi Kise

Document Information Extraction and Its Evaluation Based on Client's Relevance 35

K.C. Santosh and Abdel Belaïd

VisualDiff: Document Image Verification and Change Detection 40

Rajiv Jain and David Doermann

Oral Session 3: Document Image Processing 1

Integrating Copies Obtained from Old and New Preservation Efforts 47
Yoram Zarai, Tamar Lavee, Nachum Dershowitz, and Lior Wolf

Fast Integral MeanShift: Application to Color Segmentation of Document Images 52
Frank Lebourgeois, Fadoua Drira, Djamel Gaceb, and Jean Duong

Staff Line Detection and Removal in the Grayscale Domain 57
Ana Rebelo and Jaime S. Cardoso

Vectorization of 3D-Characters by Integral Invariant Filtering of High-Resolution Triangular Meshes 62
Hubert Mara and Susanne Krömker

Oral Session 4: Online Handwriting Recognition

An Irrelevant Variability Normalization Based Discriminative Training Approach for Online Handwritten Chinese Character Recognition 69
Jun Du and Qiang Huo

Learning-Based Candidate Segmentation Scoring for Real-Time Recognition of Online Overlaid Chinese Handwriting 74
Yan-Fei Lv, Lin-Lin Huang, Da-Han Wang, and Cheng-Lin Liu

Online Handwriting Recognition Using Levenshtein Distance Metric 79
S. Dutta Chowdhury, U. Bhattacharya, and S.K. Parui

A Semi-incremental Recognition Method for On-Line Handwritten Japanese Text 84
Cuong Tuan Nguyen, Bilan Zhu, and Masaki Nakagawa

Oral Session 5: Applications 2

The Reading-Life Log—Technologies to Recognize Texts That We Read 91
Takashi Kimura, Rong Huang, Seiichi Uchida, Masakazu Iwamura, Shinichiro Omachi, and Koichi Kise

Reading Activity Recognition Using an Off-the-Shelf EEG—Detecting Reading Activities and Distinguishing Genres of Documents 96
Kai Kunze, Yuki Shiga, Shoya Ishimaru, and Koichi Kise

Intellix—End-User Trained Information Extraction for Document Archiving 101
Daniel Schuster, Klemens Muthmann, Daniel Esser, Alexander Schill, Michael Berger, Christoph Weidling, Kamil Aliyev, and Andreas Hofmeier

A System Based on Intrinsic Features for Fraudulent Document Detection 106
Romain Bertrand, Petra Gomez-Krämer, Oriol Ramos Terrades, Patrick Franco, and Jean-Marc Ogier

Oral Session 6: Binarization

Quality Evaluation of Ancient Digitized Documents for Binarization Prediction 113
Vincent Rabeux, Nicholas Journet, Anne Vialard, and Jean-Philippe Domenger

Adaptative Smart-Binarization Method: For Images of Business Documents 118
Djamel Gaceb, Frank Lebourgeois, and Jean Duong

Color Drop-Out Binarization Method for Document Images with Color Shift 123
Minenobu Seki, Eisuke Asano, Tsukasa Yasue, Hiroto Nagayoshi, Hiroshi Shinjo, and Takeshi Nagasaki

Image Binarization for End-to-End Text Understanding in Natural Images 128
Sergey Milyaev, Olga Barinova, Tatiana Novikova, Pushmeet Kohli, and Victor Lempitsky

Poster Session 1

Figure Metadata Extraction from Digital Documents 135
Sagnik Ray Choudhury, Prasenjit Mitra, Andi Kirk, Silvia Szep, Donald Pellegrino, Sue Jones, and C. Lee Giles

Identification of Investigator Name Zones Using SVM Classifiers and Heuristic Rules 140
Jongwoo Kim, Daniel X. Le, and George R. Thoma

Enhancement of Multispectral Images of Degraded Documents by Employing Spatial Information 145
Fabian Hollaus, Melanie Gau, and Robert Sablatnig

Ukiyo-e Rakkan Retrieval System 150
Liang Li, Chulapong Panichkriangkrai, and Kozaburo Hachimura

Self Learning Classification for Degraded Document Images by Sparse Representation 155
Bolan Su, Shuangxuan Tian, Shijian Lu, Thien Anh Dinh, and Chew Lim Tan

Creating an Improved Version Using Noisy OCR from Multiple Editions 160
David Wemhoener, Ismet Zeki Yalniz, and R. Manmatha

A Comparison of Feature and Pixel-Based Methods for Recognizing Handwritten Bangla Digits 165
Olarik Surinta, Lambert Schomaker, and Marco Wiering

Part-Based Recognition of Arbitrary Fonts 170
Song Wang, Seiichi Uchida, and Marcus Liwicki

An Objective Method to Evaluate Stroke-Width Measures for Binarized Documents 175
Marte A. Ramírez-Ortegón, Volker Märgner, Raúl Rojas, and Erik Cuevas

Clustering of Symbols Using Minimal Description Length 180
Oben M. Tataw, Thanawin Rakthanmanon, and Eamonn J. Keogh

On the Evaluation of Handwritten Text Line Detection Algorithms 185
Bastien Moysset and Christopher Kermorvant

Ground-Truth Estimation in Multispectral Representation Space: Application to Degraded Document Image Binarization 190
Rachid Hedjam and Mohamed Cheriet

Degraded Digit Restoration Based on Physical Forces 195
A.N.G. Lopes Filho and C.A.B. Mello

Show-Through Cancellation and Image Enhancement by Multiresolution Contrast Processing 200
Alicia Fornés, Xavier Otazu, and Josep Lladós

Binarization of Color Historical Document Images Using Local Image Equalization and XDoG 205
Edward Roe and Carlos A.B. Mello

Automatic Enhancement and Binarization of Degraded Document Images 210
Jon Parker, Ophir Frieder, and Gideon Frieder

Ink-Bleed Reduction Using Layer Separation 215
Shrikant Baronia and Anoop Namboodiri

Application of Phase-Based Features and Denoising in Postprocessing and Binarization of Historical Document Images 220
Hossein Ziaei Nafchi, Reza Farrahi Moghaddam, and Mohamed Cheriet

A Coarse to Fine Skew Estimation Technique for Handwritten Words 225
A. Papandreou and B. Gatos

Key-Region Detection for Document Images—Application to Administrative Document Retrieval 230
Hongxing Gao, Marçal Rusiñol, Dimosthenis Karatzas, Josep Lladós, Tomokazu Sato, Masakazu Iwamura, and Koichi Kise

Modeling Local Word Spatial Configurations for Near Duplicate Document Image Retrieval 235
Li Liu, Yue Lu, Ching Y. Suen, and Jinhua Xu

Stroke-Based Character Segmentation of Low-Quality Images on Ancient Chinese Tablet 240
Xiaoqing Lu, Zhi Tang, Yan Liu, Liangcai Gao, Ting Wang, and Zhipeng Wang

A Simple Equation Region Detector for Printed Document Images in Tesseract 245
Zongyi Liu and Ray Smith

An Efficient Algorithm for Segmenting Warped Text-Lines in Document Images 250
Daniel Oliveira, Rafael Lins, Gabriel Torreão, Jian Fan, and Marcelo Thielo

Cross-Language Sensitive Words Distribution Map: A Novel Recognition-Based Document Understanding Method for Uighur and Tibetan 255
Bing Su, Xiaoqing Ding, Liangrui Peng, and Changsong Liu

Detection of Overlapped Quadrangles in Plane Geometric Figures 260
Keqiang Li, Xiaoqing Lu, Haibin Ling, Lu Liu, Tianxiao Feng, and Zhi Tang

New Approach for Symbol Recognition Combining Shape Context of Interest Points with Sparse Representation 265
Thanh Ha Do, Salvatore Tabbone, and Oriol Ramos Terrades

Improving Logo Spotting and Matching for Document Categorization by a Post-Filter Based on Homography 270
Viet Phuong Le, Muriel Visani, Cao De Tran, and Jean-Marc Ogier

Specific Comic Character Detection Using Local Feature Matching 275
Weihan Sun, Jean-Christophe Burie, Jean-Marc Ogier, and Koichi Kise

Open Vocabulary Arabic Handwriting Recognition Using Morphological Decomposition 280
Mahdi Hamdani, Amr El-Desoky Mousa, and Hermann Ney

Feature Extraction with Convolutional Neural Networks for Handwritten Word Recognition 285
Théodore Bluche, Hermann Ney, and Christopher Kermorvant

Feature Design for Offline Arabic Handwriting Recognition: Handcrafted vs Automated? 290
Youssouf Chherawala, Partha Pratim Roy, and Mohamed Cheriet

Sub-structure Learning Based Handwritten Chinese Text Recognition 295
Yuanping Zhu, Jun Sun, and Satoshi Naoi

Interactive Knowledge Learning for Ancient Images 300
Nhu-Van Nguyen, Mickael Coustaty, Alain Boucher, and Jean-Marc Ogier

WebGT: An Interactive Web-Based System for Historical Document Ground Truth Generation 305
Ofer Biller, Abedelkadir Asi, Klara Kedem, and Itshak Dinstein

Multilingual Artificial Text Detection Using a Cascade of Transforms 309
Ahsen Raza, Imran Siddiqi, Chawki Djeddi, and Abdellatif Ennaji

Efficient Word Image Retrieval Using Earth Movers Distance Embedded to Wavelets Coefficients Domain 314
Raid Saabni

A Two-Stage Approach for Word Spotting in Graphical Documents 319
Arundhati Tarafdar, Umapada Pal, Partha Pratim Roy, Nicolas Ragot, and Jean-Yves Ramel

Extraction of Spelling Variations from Language Structure for Noisy Text Correction 324
Stefan Gerdjikov, Stoyan Mihov, and Vladislav Nenchev

Automatic Chinese Text Classification Using Character-Based and Word-Based Approach 329
Xi Luo, Wataru Ohyama, Tetsushi Wakabayashi, and Fumitaka Kimura

Improving Formula Analysis with Line and Mathematics Identification 334
Mohamed Alkalai, Josef B. Baker, Volker Sorge, and Xiaoyan Lin

A Text Line Detection Method for Mathematical Formula Recognition 339
Xiaoyan Lin, Liangcai Gao, Zhi Tang, Josef Baker, Mohamed Alkalai, and Volker Sorge

Directional Discrete Cosine Transform for Handwritten Script Identification 344
Mallikarjun Hangarge, K.C. Santosh, and Rajmohan Pardeshi

Online Handwritten Cursive Word Recognition Using Segmentation-Free MRF in Combination with P2DBMN-MQDF 349
Bilan Zhu, Arti Shivram, Srirangaraj Setlur, Venu Govindaraju, and Masaki Nakagawa

ARTIST: ART-2A Driven Generation of Fuzzy Rules for Online Handwritten Gesture Recognition 354
Marta Režnáková, Lukas Tencer, and Mohamed Cheriet

A Progressive Structural Analysis Approach for Handwritten Chemical Formula Recognition 359
Peng Tang, Siu Cheung Hui, and Chi-Wing Fu

A Discriminative Approach to On-Line Handwriting Recognition Using Bi-character Models 364
S. Prum, M. Visani, A. Fischer, and J.M. Ogier

LBP Based Line-Wise Script Identification 369
Miguel A. Ferrer, Aythami Morales, and Umapada Pal

Online Signature Analysis Based on Accelerometric and Gyroscopic Pens and Legendre Series 374
Erika Griechisch, Muhammad Imran Malk, and Marcus Liwicki

Improvement of Japanese Signature Verification by Segmentation-Verification 379
Yuta Kamihira, Wataru Ohyama, Tetsushi Wakabayashi, and Fumitaka Kimura

An Improved Component Tree Based Approach to User-Intention Guided Text Extraction from Natural Scene Images 383
Lei Sun and Qiang Huo

Adaptive Scene Text Detection Based on Transferring Adaboost 388
Song Gao, Chunheng Wang, Baihua Xiao, Cunzhao Shi, Yang Zhang, Zhijian Lv, and Yanqin Shi

Rectification of Optical Characters as Transform Invariant Low-Rank Textures 393
Xin Zhang, Zhouchen Lin, Fuchun Sun, and Yi Ma

Whole is Greater than Sum of Parts: Recognizing Scene Text Words 398
Vibhor Goel, Anand Mishra, Karteek Alahari, and C.V. Jawahar

A Book Dewarping System by Boundary-Based 3D Surface Reconstruction 403
Yuan He, Pan Pan, Shufu Xie, Jun Sun, and Satoshi Naoi

Real Time Camera Phone Guidance for Compliant Document Image Acquisition without Sight 408
Michael P. Cutter and Roberto Manduchi

A New Method for Character Segmentation from Multi-oriented Video Words 413
Nabin Sharma, Palaiahnakote Shivakumara, Umapada Pal, Michael Blumenstein, and Chew Lim Tan

What Should We Be Comparing for Writer Identification? 418
Andrew J. Newell

Codebook for Writer Characterization: A Vocabulary of Patterns or a Mere Representation Space? 423
Chawki Djeddi, Imran Siddiqi, Labiba Souici-Meslati, and Abdellatif Ennaji

Text-Independent Writer Identification on Online Arabic Handwriting 428
Mariem Gargouri, Slim Kanoun, and Jean-Marc Ogier

Oral Session 7: Handwriting Recognition

Voronoi Tessellation for Effective and Efficient Handwritten Digit Classification 435
S. Impedovo, F.M. Mangini, G. Pirlo, D. Barbuzzi, and D. Impedovo

Alpha*-Approximated Delaunay Triangulation Based Descriptors for Handwritten Character Recognition 440
Octavio Razafindramanana, Frésériс Rayar, and Gilles Venturini

Rejection Schemes in Multi-class Classification—Application to Handwritten Character Recognition 445
Hubert Cecotti and Szilárd Vajda

A Comprehensive Representation Model for Handwriting Dedicated to Word Spotting 450
Peng Wang, Véronique Eglin, Christophe Garcia, Christine Largeron, and Antony McKenna

Oral Session 8: Scene Text Segmentation

Scene Text Segmentation via Inverse Rendering 457
Yahan Zhou, Jacqueline Feild, Erik Learned-Miller, and Rui Wang

Scene Character Detection by an Edge-Ray Filter 462
Rong Huang, Palaiahnakote Shivakumara, and Seiichi Uchida

Multi-script Text Extraction from Natural Scenes 467
Lluís Gómez and Dimosthenis Karatzas

On the Possibility of Structure Learning-Based Scene Character Detector 472
Yugo Terada, Rong Huang, Yaokai Feng, and Seiichi Uchida

Oral Session 9: Document Image Processing 2

Document Authentication Using Printing Technique Features and Unsupervised Anomaly Detection 479
Johann Gebhardt, Markus Goldstein, Faisal Shafait, and Andreas Dengel

Multiple Learned Dictionaries Based Clustered Sparse Coding for the Super-Resolution of Single Text Image 484
Rim Walha, Fadoua Drira, Franck Lebourgeois, Christophe Garcia, and Adel M. Alimi

Semi-synthetic Document Image Generation Using Texture Mapping on Scanned 3D Document Shapes 489
V.C. Kieu, Nicholas Journet, Muriel Visani, Rémy Mullot, and Jean Phillipe Domenger

A Probabilistic Model for Reconstruction of Torn Forensic Documents 494
Ankush Roy and Utpal Garain

Oral Session 10: Keyword Spotting 1

Fast HMM-Filler Approach for Key Word Spotting in Handwritten Documents 501
Alejandro Héctor Toselli and Enrique Vidal

Improving HMM-Based Keyword Spotting with Character Language Models 506
Andreas Fischer, Volkmar Frinken, Horst Bunke, and Ching Y. Suen

Integrating Visual and Textual Cues for Query-by-String Word Spotting 511
David Aldavert, Marçal Rusiñol, Ricardo Toledo, and Josep Lladós

Word Spotting and Regular Expression Detection in Handwritten Documents 516
Yousri Kessentini, Clément Chatelain, and Thierry Paquet

Oral Session 11: Camera-Based OCR

On Combining Multiple Segmentations in Scene Text Recognition 523
Lukáš Neumann and Jirï Matas

Automatic Ground Truth Generation of Camera Captured Documents Using Document Image Retrieval 528
Sheraz Ahmed, Koichi Kise, Masakazu Iwamura, Marcus Liwicki, and Andreas Dengel

Local Subspace Classifier with Transformation Invariance for Appearance-Based Character Recognition in Natural Images 533
Keisuke Higa and Seiji Hotta

Multiple Geometry Transform Estimation from Single Camera-Captured Text Image 538
Xin Zhang and Fuchun Sun

Oral Session 12: Writer Identification

Writer Identification and Writer Retrieval Using the Fisher Vector on Visual Vocabularies 545
Stefan Fiel and Robert Sablatnig

Writer Identification Using an Alphabet of Contour Gradient Descriptors 550
Rajiv Jain and David Doermann

Generalized Eigen Cooccurrence: Application to Palaeography 555
Ikram Moalla, Frank Lebourgeois, and Adel Alimi

CVL-DataBase: An Off-Line Database for Writer Retrieval, Writer Identification and Word Spotting 560
Florian Kleber, Stefan Fiel, Markus Diem, and Robert Sablatnig

Oral Session 13: Keyword Spotting 2

Keyword Spotting in Online Chinese Handwritten Documents with Candidate Scoring Based on Semi-CRF Model 567
Heng Zhang, Xiang-Dong Zhou, and Cheng-Lin Liu

Verification of Hierarchical Classifier Results for Handwritten Arabic Word Spotting 572
Muna Khayyat, Louisa Lam, and Ching Y. Suen

Character N-Gram Spotting on Handwritten Documents Using Weakly-Supervised Segmentation 577
Udit Roy, Naveen Sankaran, K. Pramod Sankar, and C.V. Jawahar

Part-Structured Inkball Models for One-Shot Handwritten Word Spotting 582
Nicholas R. Howe

Oral Session 14: Video and Camera-Based OCR

Recognition of Video Text through Temporal Integration 589
Trung Quy Phan, Palaiahnakote Shivakumara, Tong Lu, and Chew Lim Tan

Detection of Curved Text in Video: Quad Tree Based Method 594
P. Shivakumara, H.T. Basavaraju, D.S. Guru, and C.L. Tan

Scene Text Recognition with a Hough Forest Implicit Shape Model 599
Jae-Hyun Seok and Jin Hyung Kim

Improving Open-Vocabulary Scene Text Recognition 604
Jacqueline L. Feild and Erik G. Learned-Miller

Oral Session 15: Document Analysis and Classification

A Stream-Based Semi-supervised Active Learning Approach for Document Classification 611
Mohamed-Rafik Bouguelia, Yolande Belaïd, and Abdel Belaïd

Document Classification in a Non-stationary Environment: A One-Class SVM Approach 616
Anh Khoi Ngo Ho, Nicolas Ragot, Jean-Yves Ramel, Véronique Eglin, and Nicolas Sidere

Document Classification and Page Stream Segmentation for Digital Mailroom Applications 621
Albert Gordo, Marçal Rusiñol, Dimosthenis Karatzas, and Andrew D. Bagdanov

Continuous Partial-Order Planning for Multichannel Document Analysis: A Process-Driven Approach 626
Kristin Stamm, Marcus Liwicki, and Andreas Dengel

Poster Session 2

Label Detection and Recognition for USPTO Images Using Convolutional K-Means Feature Quantization and Ada-Boost 633
Siyu Zhu and Richard Zanibbi

Text Line Detection in Document Images: Towards a Support System for the Blind 638
Bogdan Tomoyuki Nassu, Rodrigo Minetto, and Luiz Eduardo Soares de Oliveira

Document Specific Sparse Coding for Word Retrieval 643
Ravi Shekhar and C.V. Jawahar

Achieving Linguistic Provenance via Plagiarism Detection 648
Nwokedi Idika, Harry Phan, and Mayank Varia

Detection of Cut-and-Paste in Document Images 653
Ankit Gandhi and C.V. Jawahar

Novel Sub-character HMM Models for Arabic Text Recognition 658
Irfan Ahmad, Leonard Rothacker, Gernot A. Fink, and Sabri A. Mahmoud

Statistical Modeling of the Relation between Characters and Diacritics in Lampung Script 663
Akmal Junaidi, René Grzeszick, Gernot A. Fink, and Szilárd Vajda

A Radial Neural Convolutional Layer for Multi-oriented Character Recognition 668
Hubert Cecotti and Szilárd Vajda

Sorting-Based Dynamic Classifier Ensemble Selection 673
Yan Yan, Xu-Cheng Yin, Zhi-Bin Wang, Xuwang Yin, Chun Yang, and Hong-Wei Hao

Devanagari Text Recognition: A Transcription Based Formulation 678
Naveen Sankaran, Aman Neelappa, and C.V. Jawahar

High-Performance OCR for Printed English and Fraktur Using LSTM Networks 683
Thomas M. Breuel, Adnan Ul-Hasan, Mayce Ali Al-Azawi, and Faisal Shafait

The Significance of Reading Order in Document Recognition and Its Evaluation 688
C. Clausner, S. Pletschacher, and A. Antonacopoulos

Graphical Figure Classification Using Data Fusion for Integrating Text and Image Features 693
Beibei Cheng, R. Joe Stanley, Sameer Antani, and George R. Thoma

Extraction of Serial Numbers on Bank Notes 698
Bo-Yuan Feng, Mingwu Ren, Xu-Yao Zhang, and Ching Y. Suen

Unsupervised Ensemble of Experts (EoE) Framework for Automatic Binarization of Document Images 703
Reza Farrahi Moghaddam, Fereydoun Farrahi Moghaddam, and Mohamed Cheriet

A Generic Method for Stamp Segmentation Using Part-Based Features 708
Sheraz Ahmed, Faisal Shafait, Marcus Liwicki, and Andreas Dengel

Sparse Document Image Coding for Restoration 713
Vijay Kumar, Amit Bansal, Goutam Hari Tulsiyan, Anand Mishra, Anoop Namboodiri, and C.V. Jawahar

Handwritten Line Detection via an EM Algorithm 718
Francisco Cruz and Oriol Ramos Terrades

Document Image Quality Assessment: A Brief Survey 723
Peng Ye and David Doermann

Sketch-Based Retrieval of Document Illustrations and Regions of Interest 728
Lukas Tencer, Marta Režnáková, and Mohamed Cheriet

Bringing Semantics in Word Image Retrieval 733
Praveen Krishnan and C.V. Jawahar

Automatic Detection of Pseudocodes in Scholarly Documents Using Machine Learning 738
Suppawong Tuarob, Sumit Bhatia, Prasenjit Mitra, and C. Lee Giles

Text Line Detection for Heterogeneous Documents 743
Markus Diem, Florian Kleber, and Robert Sablatnig

Towards Generic Text-Line Extraction 748
Syed Saqib Bukhari, Faisal Shafait, and Thomas M. Breuel

A Document Image Segmentation System Using Analysis of Connected Components 753
F. Zirari, A. Ennaji, S. Nicolas, and D. Mammass

Extracting Sentiment Networks from Shakespeare's Plays 758
Eric T. Nalisnick and Henry S. Baird

Automated Error Detection and Correction of Chinese Characters in Written Essays Based on Weighted Finite-State Transducer 763
Shudong Hao, Zongtian Gao, Mingqing Zhang, Yanyan Xu, Hengli Peng, Kaile Su, and Dengfeng Ke

Relation Bag-of-Features for Symbol Retrieval 768
K.C. Santosh, Laurent Wendling, and Bart Lamiroy

Constructing a Hierarchical Structure from Symbol Alphabets of Technical Line Drawings 773
Nibal Nayef and Thomas M. Breuel

Chinese Handwritten Legal Amount Recognition with HMM-Based Approach 778
Bingyu Chi and Youbin Chen

Comparative Study of HMM and BLSTM Segmentation-Free Approaches for the Recognition of Handwritten Text-Lines 783
Olivier Morillot, Laurence Likforman-Sulem, and Emmanuèle Grosicki

Category-Based Language Models for Handwriting Recognition of Marriage License Books 788
Verónica Romero and Joan Andreu Sánchez

Tamil Handwritten City Name Database Development and Recognition for Postal Automation 793
S. Thadchanamoorthy, N.D. Kodikara, H.L. Premaretne, Umapada Pal, and Fumitaka Kimura

Identification of Machine-Printed and Handwritten Words in Arabic and Latin Scripts 798
A. Saïdani, A. Kacem Echi, and A. Belaïd

A Locale Group Based Line Segmentation Approach for Non Uniform Skewed and Curved Arabic Handwritings 803
Laslo Dinges, Ayoub Al-Hamadi, and Moftah Elzobi

An Efficient Ground Truthing Tool for Binarization of Historical Manuscripts 807
Hossein Ziaei Nafchi, Seyed Morteza Ayatollahi, Reza Farrahi Moghaddam, and Mohamed Cheriet

Text Line Detection in Corrupted and Damaged Historical Manuscripts 812
Irina Rabaev, Ofer Biller, Jihad El-Sana, Klara Kedem, and Itshak Dinstein

A Pixel Labeling Approach for Historical Digitized Books 817
Maroua Mehri, Pierre Héroux, Petra Gomez-Krämer, Alain Boucher, and Rémy Mullot

Handwritten Information Extraction from Historical Census Documents 822
Thibauld Nion, Farès Menasri, Jérôme Louradour, Cédric Sibade, Thomas Retornaz, Pierre-Yves Métaireau, and Christopher Kermorvant

Segmentation-Free Keyword Spotting for Handwritten Documents Based on Heat Kernel Signature 827
Xi Zhang and Chew Lim Tan

Handwritten Musical Document Retrieval Using Music-Score Spotting 832
Rakesh Malik, Partha Pratim Roy, Umapada Pal, and Fumitaka Kimura

Greedy Search for Active Learning of OCR 837
Arpit Agarwal, Ritu Garg, and Santanu Chaudhury

GPU-Based Fast Training of Discriminative Learning Quadratic Discriminant Function for Handwritten Chinese Character Recognition 842
Ming-Ke Zhou, Fei Yin, and Cheng-Lin Liu

Mixed Thai-English Character Classification Based on Histogram of Oriented Gradient Feature 847
Teera Siriteerakul

Segmentation Based Online Word Recognition: A Conditional Random Field Driven Beam Search Strategy 852
Arti Shivram, Bilan Zhu, Srirangaraj Setlur, Masaki Nakagawa, and Venu Govindaraju

A Ballistic Stroke Representation of Online Handwriting for Recognition 857
S. Prabhu Teja and Anoop M. Namboodiri

An Empirical Comparative Study of Online Handwriting Chinese Character Recognition: Simplified vs. Traditional 862
Yan Gao, Lianwen Jin, and Weixin Yang

Word-Wise Script Identification from Video Frames 867
Nabin Sharma, Sukalpa Chanda, Umapada Pal, and Michael Blumenstein

Part-Based Automatic System in Comparison to Human Experts for Forensic Signature Verification 872
Muhammad Imran Malik, Marcus Liwicki, and Andreas Dengel

Hyperspectral Imaging for Ink Mismatch Detection 877
Zohaib Khan, Faisal Shafait, and Ajmal Mian

Finding Critical Cells in Web Tables with SRL: Trying to Uncover the Devil's Tease 882
Nicola Di Mauro, Floriana Esposito, and Stefano Ferilli

Segmenting Tables via Indexing of Value Cells by Table Headers 887
Sharad Seth and George Nagy

A Pair-Copula Based Scheme for Text Extraction from Digital Images 892
Anandarup Roy, Swapan K. Parui, and Utpal Roy

Using a Probabilistic Syllable Model to Improve Scene Text Recognition 897
Jacqueline L. Feild, Erik G. Learned-Miller, and David A. Smith

Devanagari Character Recognition in Scene Images 902
Vipin Narang, Sujoy Roy, O.V.R. Murthy, and M. Hanmandlu

Feature Representations for Scene Text Character Recognition: A Comparative Study 907
Chucai Yi, Xiaodong Yang, and Yingli Tian

Scene Text Recognition Using Co-occurrence of Histogram of Oriented Gradients 912
Shangxuan Tian, Shijian Lu, Bolan Su, and Chew Lim Tan

A Bayesian Framework for Modeling Accents in Handwriting 917
Chetan Ramaiah, Arti Shivram, and Venu Govindaraju

Feature Selection for Forensic Handwriting Identification 922
Aline Maria M.M. Amaral, Cinthia Obladen de Almendra Freitas, and Flávio Bortolozzi

Alternatives for Page Skew Compensation in Writer Identification 927
Jin Chen and Daniel Lopresti

Oral Session 16: Handwritten Text Recognition 1

Improvements in RWTH's System for Off-Line Handwriting Recognition 935
Michal, Kozielski, Patrick Doetsch, and Hermann Ney

Minimum Risk Training for Handwritten Chinese/Japanese Text Recognition Using Semi-Markov Conditional Random Fields 940
Xiang-Dong Zhou, Feng Tian, Cheng-Lin Liu, and Hong-An Wang

A Hidden Markov Model-Based Approach with an Adaptive Threshold Model for Off-Line Arabic Handwriting Recognition 945
Moftah Elzobi, Ayoub Al-Hamadi, Laslo Dings, Mahmoud Elmezain, and Anwar Saeed

Oral Session 17: Layout Analysis

Unified Performance Evaluation for OCR Zoning: Calculating Page Segmentation's Score, That Includes Text Zones, Tables and Non-text Objects 953
Dmitry Deryagin

Hybrid Page Segmentation with Efficient Whitespace Rectangles Extraction and Grouping 958
Kai Chen, Fei Yin, and Cheng-Lin Liu

A Model Based Framework for Table Processing in Degraded Document Images 963
Zhixin Shi, Srirangaraj Setlur, and Venu Govindaraju

Oral Session 18: Signature Verification

FREAK for Real Time Forensic Signature Verification 971
Muhammad Imran Malik, Sheraz Ahmed, Marcus Liwicki, and Andreas Dengel

Large-Scale Signature Matching Using Multi-stage Hashing 976
Xianzhi Du, Wael Abdalmageed, and David Doermann

Can Signature Biometrics Address Both Identification and Verification Problems? 981
Salman H. Khan, Zeashan Khan, and Faisal Shafait

Oral Session 19: Handwritten Text Recognition 2

Using the Web to Create Dynamic Dictionaries in Handwritten Out-of-Vocabulary Word Recognition 989
Cristina Oprean, Laurence Likforman-Sulem, Adrian Popescu, and Chafic Mokbel

Detecting OOV Names in Arabic Handwritten Data 994
Jinying Chen, Rohit Prasad, Huaigu Cao, and Premkumar Natarajan

A Stroke Order Verification Method for On-Line Handwritten Chinese Characters Based on Tempo-spatial Consistency Analysis 999
Rongsha Li, Liangrui Peng, Endong Xun, and Nan Wein

Oral Session 20: Graphics Recognition 1

Graphics Extraction from Heterogeneous Online Documents with Hierarchical Random Fields 1007
Adrien Delaye and Cheng-Lin Liu

Classification of On-Line Mathematical Symbols with Hybrid Features and Recurrent Neural Networks 1012
Francisco Álvaro, Joan-Andreu Sánchez, and José-Miguel Benedí

Using Confusion Reject to Improve (User and) System (Cross) Learning of Gesture Commands 1017
Manuel Bouillon, Peiyu Li, Eric Anquetil, and Grégoire Richard

Deformable HOG-Based Shape Descriptor 1022
Jon Almazán, Alicia Fornés, and Ernest Valveny

Oral Session 21: Historical Documents

Text Line Extraction Using DMLP Classifiers for Historical Manuscripts 1029
Micheal Baechler, Marcus Liwicki, and Rolf Ingold

Exploiting Stroke Orientation for CRF Based Binarization of Historical Documents 1034
Xujun Peng, Huaigu Cao, Krishna Subramanian, Rohit Prasad, and Prem Natarajan

Spot It! Finding Words and Patterns in Historical Documents 1039
Vladislavs Dovgalecs, Alexandre Burnett, Pierrick Tranouez, Stéphane Nicolas, and Laurent Heutte

Toponym Recognition in Historical Maps by Gazetteer Alignment 1044
Jerod Weinman

Oral Session 22: Character Recognition 2

Style Consistent Perturbation for Handwritten Chinese Character Recognition 1051
Fei Yin, Ming-Ke Zhou, Qiu-Feng Wang, and Cheng-Lin Liu

Similar Pattern Discriminant Analysis for Improving Chinese Character Recognition Accuracy 1056
Yanwei Wang, Changsong Liu, and Xiaoqing Ding

Offline Printed Urdu Nastaleeq Script Recognition with Bidirectional LSTM Networks 1061
Adnan Ul-Hasan, Saad Bin Ahmed, Faisal Rashid, Faisal Shafait, and Thomas M. Breuel

Moment-Based Character-Normalization Methods Using a Contour Image Combined with an Original Image 1066
Toshinori Miyoshi, Takeshi Nagasaki, and Hiroshi Shinjo

Oral Session 23: Graphics Recognition 2

User-Centered Design of an Interactive Off-Line Handwritten Architectural Floor Plan Recognition 1073
Sylvain Fleury, Achraf Ghorbel, Aurélie Lemaitre, Eric Anquetil, and Eric Jamet

Near Convex Region Adjacency Graph and Approximate Neighborhood String Matching for Symbol Spotting in Graphical Documents 1078
Anjan Dutta, Josep Lladós, Horst Bunke, and Umapada Pal

Robust Symbol Localization Based on Junction Features and Efficient Geometry Consistency Checking 1083
The Anh Pham, Mathieu Delalandre, Sabine Barrat, and Jean-Yves Ramel

Discriminative Weighting and Subspace Learning for Ensemble Symbol Recognition 1088
Feng Su and Tong Lu

Poster Session 3

A Novel Multi-view Object Class Detection Framework for Document Image Content Analysis 1095
Weichong Yin, Tong Lu, and Feng Su

Field Extraction from Administrative Documents by Incremental Structural Templates 1100
Marçal Rusiñol, Tayeb Benkhelfallah, and Vincent Poulain d'Andecy

Analysis of Topographic Maps for Recreational Purposes Using Decision Trees 1105
Richard Kirby and Thomas C. Henderson

Mental Workload Classification via Online Writing Features 1110
Kun Yu, Julien Epps, and Fang Chen

A New Method for Discriminating Printers Based on Contours Qualities of Printed Characters Using Wavelet Decomposition 1115
Takeshi Furukawa

Holistic Arabic Whole Word Recognition Using HMM and Block-Based DCT 1120
Abdulwahab Krayem, Nasser Sherkat, Lindsay Evett, and Taha Osman

Search Space Reduction for Holistic Ligature Recognition in Urdu Nastalique Script 1125
Akram El-Korashy and Faisal Shafait

Ligature Segmentation for Urdu OCR 1130
Gurpreet Singh Lehal

Error Detection in Highly Inflectional Languages 1135
Naveen Sankaran and C.V. Jawahar

An Anytime Algorithm for Camera-Based Character Recognition 1140
Takuya Kobayashi, Masakazu Iwamura, Takahiro Matsuda, and Koichi Kise

eBDtheque: A Representative Database of Comics 1145
Clément Guérin, Christophe Rigaud, Antoine Mercier, Farid Ammar-Boudjelal, Karell Bertet, Alain Bouju, Jean-Christophe Burie, Georges Louis, Jean-Marc Ogier, and Arnaud Revel

Script Identification of Pre-segmented Multi-font Characters and Digits 1150
Rajneesh Rani, Renu Dhir, and Gurpreet Singh Lehal

Triangular Mesh Based Stroke Segmentation for Chinese Calligraphy 1155
Xiaoqing Wang, Xiaohui Liang, Linjia Sun, and Min Liu

Detecting Main Body Size in Document Images 1160
Paraskevas Diamantatos, Vasileios Verras, and Ergina Kavallieratou

Discrete CRF Based Combination Framework for Document Image Binarization 1165
David Hebert, Stephane Nicolas, and Thierry Paquet

Automatic Selection of Binarization Method for Robust OCR 1170
T. Chattopadhyay, V. Ramu Reddy, and Utpal Garain

An OCR System with OCRopus for Scientific Documents Containing Mathematical Formulas 1175
F. Furukori, S. Yamazaki, T. Miyagishi, K. Shirai, and M. Okamoto

Segmenting Handwritten Math Symbols Using AdaBoost and Multi-scale Shape Context Features 1180
Lei Hu and Richard Zanibbi

Learning to Detect Tables in Scanned Document Images Using Line Information 1185
T. Kasar, P. Barlas, S. Adam, C. Chatelain, and T. Paquet

Unsupervised Speech Text Localization in Comic Images 1190
Luyuan Li, Yongtao Wang, Zhi Tang, Xiaoqing Lu, and Liangcai Gao

A Fast Word Retrieval Technique Based on Kernelized Locality Sensitive Hashing 1195
Tanmoy Mondal, Nicolas Ragot, Jean-Yves Ramel, and Umapada Pal

Multi-modal Information Integration for Document Retrieval 1200
Ehtesham Hassan, Santanu Chaudhury, and M. Gopal

Table of Contents Recognition and Extraction for Heterogeneous Book Documents 1205
Zhaohui Wu, Prasenjit Mitra, and C. Lee Giles

Fusion of Statistical and Structural Information for Flowchart Recognition 1210
Cérès Carton, Aurélie Lemaitre, and Bertrand Coüasnon

Modeling Flowchart Structure Recognition as a Max-Sum Problem 1215
Martin Bresler, Daniel Prùša, and Václav Hlavác

Evaluation of SVM, MLP and GMM Classifiers for Layout Analysis of Historical Documents 1220
Hao Wei, Micheal Baechler, Fouad Slimane, and Rolf Ingold

Unsupervised Classification of Structurally Similar Document Images 1225
Jayant Kumar and David Doermann

Text Line Extraction Method Using Domain-Based Active Contour Model 1230
Yusuke Itani, Takashi Hirano, and Jun Ishii

Logo Detection Using Painting Based Representation and Probability Features 1235
Alireza Alaei, Mathieu Delalandre, and Nathalie Girard

An Active Contour Model for Speech Balloon Detection in Comics 1240
Christophe Rigaud, Jean-Christophe Burie, Jean-Marc Ogier, Dimosthenis Karatzas, and Joost Van De Weijer

Unsupervised Wall Detector in Architectural Floor Plans 1245
Lluís-Pere de las Heras, David Fernández, Ernest Valveny, Josep Lladós, and Gemma Sánchez

A Novel Baseline-independent Feature Set for Arabic Handwriting Recognition 1250
Bing Su, Xiaoqing Ding, Liangrui Peng, and Changsong Liu

Human Evaluation of the Transcription Process of a Marriage License Book 1255
Verónica Romero and Joan Andreu Sánchez

An Approach for Arabic Handwriting Synthesis Based on Active Shape Models 1260
Laslo Dinges, Ayoub Al-Hamadi, and Moftah Elzobi

Lexicon Reduction Using Segment Descriptors for Arabic Handwriting Recognition 1265
Mohammad Tanvir Parvez and Sabri A. Mahmoud

Stabilize Sequence Learning with Recurrent Neural Networks by Forced Alignment 1270
Marc-Peter Schambach and Sheikh Faisal Rashid

Exploring MPE/MWE Training for Chinese Handwriting Recognition 1275
Tonghua Su, Peijun Ma, Tong Wei, Shu Liu, and Shengchun Deng

Interactive Off-Line Handwritten Text Transcription Using On-Line Handwritten Text as Feedback 1280
Daniel Martín-Albo, Verónica Romero, and Enrique Vidal

Character Shape Restoration of Binarized Historical Documents by Smoothing via Geodesic Morphology 1285
K. Shirai, Y. Endo, A. Kitadai, S. Inoue, N. Kurushima, H. Baba, A. Watanabe, and M. Nakagawa

A Binarization-Free Clustering Approach to Segment Curved Text Lines in Historical Manuscripts 1290
Angelika Garz, Andreas Fischer, Horst Bunke, and Rolf Ingold

Using Harris Corners for the Retrieval of Graphs in Historical Manuscripts 1295
Rainer Herzog, Arved Solth, and Bernd Neumann

On Evaluation of Segmentation-Free Word Spotting Approaches without Hard Decisions 1300
Werner Pantke, Volker Märgner, and Tim Fingscheidt

Bag-of-Features HMMs for Segmentation-Free Word Spotting in Handwritten Documents 1305
Leonard Rothacker, Marçal Rusiñol, and Gernot A. Fink

OCR-Free Transcript Alignment 1310
Tal Hassner, Lior Wolf, and Nachum Dershowitz

An Empirical Evaluation of Supervised Dimensionality Reduction for Recognition 1315
Guoqiang Zhong, Youssouf Chherawala, and Mohamed Cheriet

Bayesian Network Structure Learning and Inference Methods for Handwriting 1320
Mukta Puri, Sargur N. Srihari, and Yi Tang

Invariants Extraction Method Applied in an Omni-language Old Document Navigating System 1325
Quang Anh Bui, Muriel Visani, and Rémy Mullot

A Multi-stroke Dynamic Time Warping Distance Based on A* Optimization 1330
Jinpeng Li, Harold Mouchere, Christian Viard-Gaudin, and Zhaoxin Chen

A System for Bangla Online Handwritten Text 1335
Nilanjana Bhattacharya, Umapada Pal, and Fumitaka Kimura

Semi-automatic Tibetan Component Annotation from Online Handwritten Tibetan Character Database by Optimizing Segmentation Hypotheses 1340
Long-Long Ma and Jian Wu

Offline Signature Verification Using Real Adaboost Classifier Combination of Pseudo-dynamic Features 1345
Juan Hu and Youbin Chen

A Color-Based Model to Determine the Age of Documents for Forensic Purposes 1350
Ricardo da Silva Barboza, Rafael Dueire Lins, and Darlisson Marinho de Jesus

A Novel Multi-oriented Chinese Text Extraction Approach from Videos 1355
Yang Liu, Yonghong Song, Yuanlin Zhang, and Quan Meng

Scene Character Reconstruction through Medial Axis 1360
Shangxuan Tian, Palaiahnakote Shivakumara, Trung Quy Phan, and Chew Lim Tan

Automatic Labeling for Scene Text Database 1365
Masakazu Iwamura, Masaki Tsukada, and Koichi Kise

Text Detection in Natural Images Using Bio-inspired Models 1370
Konstantinos Zagoris and Ioannis Pratikakis

Natural Scene Text Detection with Multi-channel Connected Component Segmentation 1375
Xiaobing Wang, Yonghong Song, and Yuanlin Zhang

Scene Text Localization Using Gradient Local Correlation 1380
Bo Bai, Fei Yin, and Cheng Lin Liu

Discriminating Features for Writer Identification 1385
Zachary A. Daniels and Henry S. Baird

Most Discriminative Primitive Selection for Identity Determination Using Handwritten Devanagari Script 1390
Nivedita Yadav, Santanu Chaudhury, and Prem Kalra

Competition Papers

ICDAR 2013 Competition on Writer Identification 1397
G. Louloudis, B. Gatos, N. Stamatopoulos, and A. Papandreou

ICDAR 2013 Handwriting Segmentation Contest 1402
Nikolaos Stamatopoulos, Basilis Gatos, Georgios Louloudis, Umapada Pal, and Alireza Alaei

ICDAR 2013 Music Scores Competition: Staff Removal 1407
Muriel Visaniy, V.C Kieu, Alicia Fornés, and Nicholas Journet

ICDAR 2013 Competition on Handwriting Stroke Recovery from Offline Data 1412
Abdelâali Hassaïne, Somaya Al Maadeed, and Ahmed Bouridane

ICDAR 2013 Competition on Gender Prediction from Handwriting 1417
Abdelâali Hassaïne, Somaya Al Maadeed, Jihad Aljaam, and Ali Jaoua

ICDAR 2013 Competition on Handwritten Digit Recognition (HDRC 2013) 1422
Markus Diem, Stefan Fiel, Angelika Garz, Manuel Keglevic, Florian Kleber, and Robert Sablatnig

ICDAR 2013 CROHME: Third International Competition on Recognition of Online Handwritten Mathematical Expressions 1428
Harold Mouchère, Christian Viard-Gaudin, Richard Zanibbi, Utpal Garain, Dae Hwan Kim, and Jin Hyung Kim

ICDAR2013 Competition on Multi-font and Multi-size Digitally Represented Arabic Text 1433
Fouad Slimane, Slim Kanoun, Haikal El Abed, Adel M. Alimi, Rolf Ingold, and Jean Hennebert

ICDAR 2013 Competition on Book Structure Extraction 1438
Antoine Doucet, Gabriella Kazai, Sebastian Colutto, and Günter Mühlberger

ICDAR 2013 Document Image Skew Estimation Contest (DISEC 2013) 1444
A. Papandreou, B. Gatos, G. Louloudis, and N. Stamatopoulos

ICDAR 2013 Table Competition 1449
Max Göbel, Tamir Hassan, Ermelinda Oro, and Giorgio Orsi

ICDAR 2013 Competition on Historical Newspaper Layout Analysis (HNLA 2013) 1454
A. Antonacopoulos, C. Clausner, C. Papadopoulos, and S. Pletschacher

ICDAR 2013 Competition on Historical Book Recognition (HBR 2013) 1459
A. Antonacopoulos, C. Clausner, C. Papadopoulos, and S. Pletschacher

ICDAR 2013 Chinese Handwriting Recognition Competition 1464
Fei Yin, Qiu-Feng Wang, Xu-Yao Zhang, and Cheng-Lin Liu

ICDAR 2013 Document Image Binarization Contest (DIBCO 2013) 1471
Ioannis Pratikakis, Basilis Gatos, and Konstantinos Ntirogiannis

ICDAR 2013 Competitions on Signature Verification and Writer Identification for On- and Offline Skilled Forgeries (SigWiComp 2013) 1477
Muhammad Imran Malik, Marcus Liwicki, Linda Alewijnse, Wataru Ohyama, Michael Blumenstein, and Bryan Found

ICDAR 2013 Robust Reading Competition 1484
Dimosthenis Karatzas, Faisal Shafait, Seiichi Uchida, Masakazu Iwamura, Lluis Gomez i Bigorda, Sergi Robles Mestre, Joan Mas, David Fernandez Mota, Jon Almazàn Almazàn, and Lluís Pere de las Heras

Author Index 1494