# Table of Contents

Automatic Extraction of Correlation-Entropy Features for Text Document Analysis Directly in Run-Length Compressed Domain .......................................................... 1  
*Mohammed Javed, P Nagabhushan and Bidyut Baran Chaudhuri*

A Polar Stroke Descriptor for Classification of Historical Documents ......................... 6  
*Sheng He and Lambert Schomaker*

Solving Substitution Ciphers for OCR with a Semi-supervised Hidden Markov Model ........ 11  
*Erik Scharwächter and Stephan Vogel*

Co-occurrence Matrix of Oriented Gradients for Word Script and Nature Identification ........ 16  
*Asma Saidani, Afef Kacem and Abdel Belaid*

A Recognition based Approach for Segmenting Touching Components in Arabic Manuscripts .... 21  
*Nabil Aouadi, Afef Kacem and Abdel Belaid*

Towards a SignWriting Recognition System ............................................................... 26  
*Diego Stiehl, Luiz S. Oliveira, Cayley Guimarães and Alceu S. Britto Jr*

Combination of Multiple Aligned Recognition Outputs using WFST and LSTM ................ 31  
*Mayce Al Azawi, Marcus Liwicki and Thomas M Breuel*

Class-Adaptive Zoning Methods for Recognizing Handwritten Digits and Characters ........... 36  
*Donato Impedovo and Giuseppe Pirlo*

Keyword spotting in handwritten documents based on a generic text line HMM and a SVM verification ............................................................. 41  
*Yousri Kessentini and Thierry Paquet*

Online Handwritten Tibetan Syllable Recognition based on Component Segmentation Method ... 46  
*Long-Long Ma and Jian Wu*

Arabic Handwritten Words Off-line Recognition ..................................................... 51  
*Akrám Khemiri, Afef Kacem, Abdel Belaid and Mourad Elloumi*

Binarizing Complex Scanned Documents ..................................................................... 56  
*Rafael Lins, Gabriel Silva and Marcos Martins de Almeida*

Recognition Confidence Analysis of Handwritten Chinese Character with CNN ............... 61  
*Meijun He, Shuye Zhang, Huiyun Mao and Lianwen Jin*

Multi-Strategy Tracking Based Text Detection in Scene Videos .................................. 66  
*Ze-Yu Zuo, Shu Tian, Wei-Yi Pei and Xu-Cheng Yin*

Recognition of Urdu Ligatures - A Holistic Approach .............................................. 71  
*Israr Uddin Khattak, Imran Siddiqi, Shehzad Khalid and Chawki Djeddi*

Cost-sensitive MQDF Classifier for Handwritten Chinese Address Recognition ............... 76  
*Shujing Lu, Xiaohua Wei and Yue Lu*

Framewise and CTC Training of Neural Networks for Handwriting Recognition ............... 81  
*Théodore Bluche, Hermann Ney, Jérôme Louradour and Christopher Kermorvant*

The LIMSI Handwriting Recognition System for the HTRItS 2014 Contest .................... 86  
*Théodore Bluche, Hermann Ney and Christopher Kermorvant*
<table>
<thead>
<tr>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Text-independent Writer Identification Using SIFT Descriptor and Contour-directional Feature</td>
<td>91</td>
</tr>
<tr>
<td>Yu-Jie Xiong, Ying Wen, Patrick S.P. Wang and Yue Lu</td>
<td></td>
</tr>
<tr>
<td>Multi-font Printed Chinese Character Recognition using Multi-pooling Convolutional Neural Network</td>
<td>96</td>
</tr>
<tr>
<td>Zhuoyao Zhong, Lianwen Jin and Ziyong Feng</td>
<td></td>
</tr>
<tr>
<td>Similarity-based Regularization for Semi-Supervised Learning for Handwritten Digit Classification</td>
<td>101</td>
</tr>
<tr>
<td>Donato Barbuzzi, Giuseppe Pirlo, Seiichi Uchida, Volkmar Frinken and Donato Impedovo</td>
<td></td>
</tr>
<tr>
<td>Text Detection in Nature Scene Images Using Two-stage Nontext Filtering</td>
<td>106</td>
</tr>
<tr>
<td>Qingqing Wang, Yue Lu and Shiliang Sun</td>
<td></td>
</tr>
<tr>
<td>Classification of Forms With Similar Layouts by Using the Mixed Gaussian Weighted Mask (MGWM)</td>
<td>111</td>
</tr>
<tr>
<td>Simeng Wang, Liangcai Gao and Yuehan Wang</td>
<td></td>
</tr>
<tr>
<td>A Structural Signature Based on Texture for Digitized Historical Book Page Categorization</td>
<td>116</td>
</tr>
<tr>
<td>Maroua Mehri, Pierre Héroux, Julien Lerouge, Petra Gomez-Krümer and Rémy Mullot</td>
<td></td>
</tr>
<tr>
<td>A Multiple Instances Approach to Improving Keyword Spotting on Historical Mongolian Document Images</td>
<td>121</td>
</tr>
<tr>
<td>Hongxi Wei, Guanglai Gao and Xiangdong Su</td>
<td></td>
</tr>
<tr>
<td>Combining Handwriting and Speech Recognition for Transcribing Historical Handwritten Documents</td>
<td>126</td>
</tr>
<tr>
<td>Emilio Granell and Carlos-D Martínez-Hinarejos</td>
<td></td>
</tr>
<tr>
<td>Seamless Stitching with Shape Deformation for Historical Document Images</td>
<td>131</td>
</tr>
<tr>
<td>Wei Liu, Wei Fan, Li Chen, Jun Sun and Naoli Satoshi</td>
<td></td>
</tr>
<tr>
<td>An Open Source Testing Tool for Evaluating Handwriting Input Methods</td>
<td>136</td>
</tr>
<tr>
<td>Liquan Qiu, Lianwen Jin, Ruifen Dai, Yuxiang Zhang and Lei Li</td>
<td></td>
</tr>
<tr>
<td>Using Multiple Sequence Alignment and Statistical Language Model to Integrate Multiple Chinese Address Recognition Outputs</td>
<td>151</td>
</tr>
<tr>
<td>Shengchang Chen, Shujing Lu, Ying Wen and Yue Lu</td>
<td></td>
</tr>
<tr>
<td>Document Analysis by a Mobile Robot for Autonomous Indoor Navigation</td>
<td>156</td>
</tr>
<tr>
<td>Dalia Marcela Rojas Castro, Arnaud Revel and Michel Ménard</td>
<td></td>
</tr>
<tr>
<td>HoG based Two-Directional Dynamic Time Wrapping for Handwritten Word Spotting</td>
<td>161</td>
</tr>
<tr>
<td>Shunyi Yao, Ying Wen and Yue Lu</td>
<td></td>
</tr>
<tr>
<td>Evaluation of Neural Network Language Models in Handwritten Chinese Text Recognition</td>
<td>166</td>
</tr>
<tr>
<td>Yi-Chao Wu, Fei Yin and Cheng-Lin Liu</td>
<td></td>
</tr>
<tr>
<td>Segmentation-free Handwritten Chinese Text Recognition with LSTM-RNN</td>
<td>171</td>
</tr>
<tr>
<td>Ronaldo Messina and Jérôme Louradour</td>
<td></td>
</tr>
<tr>
<td>Document Image Quality Assessment based on Improved Gradient Magnitude Similarity Deviation</td>
<td>176</td>
</tr>
<tr>
<td>Alireza Alaei, Donatello Conte and Romain Raveaux</td>
<td></td>
</tr>
</tbody>
</table>
Proceedings of 13th International Conference on Document Analysis and Recognition (ICDAR’2015)

Author Identification by Automatic Learning ................................................................. 181
    Jordan Frery, Christine Largeron and Mihaela Juganaru-Mathieu

A Syntax Directed System for the Recognition of Printed Arabic Mathematical Formulas ............ 186
    Kawther Khazri, Afef Kacem and Abdel Belaid

Text Line Extraction in Document Images ...................................................................... 191
    Liuan Wang, Hiroshi Tanaka, Wei Fan, Jun Sun and Satoshi Naoi

An Improved Artificial Immune Recognition System for Off-line Handwritten Signature
Veriﬁcation ......................................................................................................................... 196
    Yasmine Serdouk, Hassiba Nemmour and Youcef Chibani

Benchmarking discriminative approaches for word spotting in handwritten documents .......... 201
    Gautier Bideault, Luc Mioleut, Clément Chatelain and Thierry Paquet

Object Proposals for Text Extraction in the Wild .............................................................. 206
    Lluis Gomez and Dimosthenis Karatzas

Topological Simpliﬁcation of Electrical Circuits by Super-component Analysis ......................... 211
    Paramita De, Sekhar Mandal, Partha Bhowmick and Bhabatosh Chanda

A Direct Approach for Word and Character Segmentation in Run-Length Compressed
Documents with an Application to Word Spotting ........................................................... 216
    Mohammed Javed, P Nagabhushan and Bidyut Baran Chaudhuri

Using histogram representation and Earth Mover’s Distance as an evaluation tool for text detection . 221
    Stefania Calarasanu, Jonathan Fabrizio and Séverine Dubuisson

A Bottom-up Method Using Texture Features and a Graph-based Representation for Lettrine
Recognition and Classiﬁcation ........................................................................................... 226
    Maroua Mehri, Petra Gomez-Krammer, Pierre Héroux, Mickaël Cousiat, Julien Lerouge and Rémy Mullot

Machine-Readable Region Identiﬁcation from Partially Blurred Document Images .................... 231
    Qinwen Wang, Yixue Wang, Chenyang Wang, Jufeng Yang, Tao Li and Kai Wang

Stretching Deep Architectures for Text Recognition ................................................................ 236
    Yuchen Zheng, Yajuan Cai, Guoqiang Zhong, Chherawala Youssouf, Yaxin Shi and Junyu Dong

Robust Score Normalization for DTW-Based On-Line Signature Veriﬁcation ......................... 241
    Andreas Fischer, Moises Diaz, Réjean Plamondon and Miguel A. Ferrer

A proposal of a document image reading-life log based on document image retrieval and
eyetracking ......................................................................................................................... 246
    Olivier Augereau, Koichi Kise and Kensuke Hoshika

The Eye as the Window of the Language Ability: Estimation of English Skills by Analyzing Eye
Movement While Reading Documents .............................................................................. 251
    Kazuyo Yoshimura, Kai Kunze and Koichi Kise

Bagging by Design for Continuous Handwriting Recognition Using Multi-Objective Particle
Swarm Optimization ........................................................................................................... 256
    Mahdi Hamdani, Patrick Doetsch and Hermann Ney
Investigation of Segmental Conditional Random Fields for Large Vocabulary Handwriting Recognition ................................................................. 261
  Mahdi Hamdani, Mahaboob Ali Basha Shaik, Patrick Doetsch and Hermann Ney

Aligning transcript of historical documents using energy minimization .................................................. 266
  Rafi Cohen, Irina Rabaev, Ishak Dinstein, Jihad El-Sana and Klara Kedem

Extracting Structured Data from Unstructured Document with Incomplete Resources ....................... 271
  Hervé Déjean

Classifier Self-Assessment: Active Learning and Active Noise Correction for Document Classification .......................................................... 276
  Dominik Henter, Armin Stahl, Markus Ebbecke and Michael Gillmann

Goal-Oriented Performance Evaluation Methodology for Page Segmentation Techniques ............... 281
  Nikolaos Stamatopoulos, Georgios Louloudis and Basilis Gatos

Improving Sigma-Lognormal Parameter Extraction ................................................................................. 286
  Daniel Martín-Albo, Réjean Plamondon and Enrique Vidal

Efficient Text Localization in Born-Digital Images by Local Contrast-Based Segmentation ........... 291
  Kai Chen, Fei Yin, Amir Hussain and Cheng-Lin Liu

Table Structure Extraction in Handwritten Chemistry Documents ...................................................... 296
  Nabil Ghanmi and Abdel Belaïd

Semantic Label and Structure Model based approach for Entity Recognition in Database Context .... 301
  Nihel Kooli and Abdel Belaïd

Preselection of Support Vector Candidates by Relative Neighborhood Graph for Large-Scale Character Recognition .................................................................................................................. 306
  Masanori Goto, Ryosuke Ishida and Seiichi Uchida

A Subtractive Clustering Scheme for Text-Independent Online Writer Identification ...................... 311
  Gautam Singh and Suresh Sundaram

Automatic Annotation Extension and Classification of Documents Using a Probabilistic Graphical Model ........................................................................................................................................... 316
  Abdessalem Bouzaïeni, Sabine Barrat and Salvatore Tabbone

A Multiple-Expert Binarization Framework for Multispectral Images ............................................. 321
  Reza Farrahi Moghaddam and Mohamed Cheriet

Character Retrieval of Vectorized Cuneiform Script ............................................................................ 326
  Bartosz Bogacz, Michael Gertz and Hubert Mara

Robust Text Segmentation using Graph Cut .............................................................................................. 331
  Shangxuan Tian, Shijian Lu, Bolan Su and Chew Lim Tan

Isolated Character Recognition using Projections of Oriented Gradients ........................................ 336
  George Retsinas, Basilis Gatos, Nikolaos Stamatopoulos and Georgios Louloudis

A combined Convolutional Neural Network and Dynamic Programming approach for text line normalization ......................................................................................................................... 341
  Joan Pastor-Pellicer, Salvador España-Boquera, María Jose Castro-Bleda and Francisco Zamora-Martínez
Proceedings of 13th International Conference on Document Analysis and Recognition (ICDAR’2015)

A segmentation free Word Spotting for handwritten documents ................................. 346
Nicole Vincent, Adam Ghorbel and Jean-Marc Ogier

Speech balloon and speaker association for comics and manga understanding ............ 351
Christophe Rigaud, Nam Le Thanh, Jean-Christophe Burie, Jean-Marc Ogier, Motoi Iwata, Eiki Imazu and Koichi Kise

Graph matching versus bag of graph for lettrines recognition ................................. 356
Mickael Coustaty and Jean-Marc Ogier

Detecting Dense Foreground Stripes in Arabic Handwriting for Accurate Baseline Positioning .... 361
Felix Stahlberg and Stephan Vogel

Document Skew Detection Based on Hough Space Derivatives .................................. 366
Felix Stahlberg and Stephan Vogel

Efficient Estimation of Character Normal Direction for camera-based OCR ................. 371
Kanta Kuramoto, Wataru Ohyama, Tetsushi Wakabayashi and Fumitaka Kimura

Content-Independent Font Recognition on a Single Chinese Character using Sparse Representation 376
Weikang Song, Zhouhui Lian, Yingmin Tang and Jianguo Xiao

Inkball Models for Character Localization and Out-of-Vocabulary Word Spotting ............ 381
Nicholas Howe

Segmented Handwritten Text Recognition with Recurrent Neural Network Classifiers .......... 386
Bolan Su, Xi Zhang, Shijian Lu and Chew Lim Tan

A New Method based on Bag of Filters for Character Recognition in Scene Images by Learning ... 391
Qisu Li, Tong Lu, Palaiahnakote Shivakumara, Umapada Pal and Chew Lim Tan

Scene Character Recognition using Markov Random Field ..................................... 396
Xiaolong Liu and Tong Lu

Study of Two Zone-based Features for Online Bengali and Devanagari Character Recognition .... 401
Rajib Ghosh and Partha Pratim Roy

Building Handwriting Recognizers by Leveraging Skeletons of Both Offline and Online Samples ... 406
Xiong Zhang, Min Wang, Lijuan Wang, Qiang Huo and Haifeng Li

A Context-Sensitive-Chunk BPTT Approach to Training Deep LSTM/BLSTM Recurrent Neural Networks for Offline Handwriting Recognition ......................... 411
Kai Chen, Zhi-Jie Yan and Qiang Huo

A Fast Color Barcode Detection Method through Cross Identification on Mobile Platforms .... 416
Yu Zhang and Tong Lu

Lexicon-Driven Recognition of One-Stroke Character Strings in Visual Gesture ............ 421
Fei Yin, Pai-Pai Liu, Linlin Huang and Cheng-Lin Liu

Scene Text Detection with Robust Character Candidate Extraction Method ................. 426
Myung-Chul Sung, Bongjin Jun, Hojin Cho and Daijin Kim

Reconstruction Combined Training for Convolutional Neural Networks on Character Recognition ... 431
Li Chen, Song Wang, Wei Fan, Jun Sun and Satoshi Naoi

Deep Learning Based Language and Orientation Recognition in Document Analysis ........ 436
Li Chen, Song Wang, Wei Fan, Jun Sun and Satoshi Naoi
Exploring the World of Fonts for Discovering the Most Standard Fonts and the Missing Fonts . . . . 441  
Seiichi Uchida, Yuji Egashira and Kota Sato

Learning Non-Markovian Constraints for Handwriting Recognition .......................... 446  
Ryosuke Kakisako, Seiichi Uchida and Volkmar Frinken

Arabic handwritten document preprocessing and recognition ........................................ 451  
Edgard Chammas, Chafic Mokbel and Laurence Likforman-Sulem

Paragraph text segmentation into lines with Recurrent Neural Networks .................... 456  
Bastien Moysset, Christopher Kermorvant, Christian Wolf and Jérôme Louradour

A Study on Effects of Implicit and Explicit Language Model Information for DBLSTM-CTC  
Based Handwriting Recognition ................................................................. 461  
Qi Liu, Lijuan Wang and Qiang Huo

BLSTM-based handwritten text recognition using Web resources ............................. 466  
Cristina Oprean, Laurence Likforman-Sulem, Chafic Mokbel and Adrian Popescu

Subspace method with multi scale wavelet for recognition of printer property ............... 471  
Takeshi Furukawa

Training an Arabic handwriting recognizer without a handwritten training data set .......... 476  
Irfan Ahmad and Gernot Fink

Novel Line Verification for Multiple Instance Focused Retrieval in Document Collections .... 481  
Hongxing Gao, Marçal Rusiñol, Dimosthenis Karatzas, Josep Llados, Rajiv Jain and David  
Doermann

Writer Identification from Offline Isolated Bangla Characters and Numerals ................ 486  
Chandranath Adak and Bidyut B. Chaudhuri

Generation of synthetic training data for handwritten Indic script recognition .......... 491  
Shivansh Gaur, Siddhant Sonkar and Partha Roy

Localized Forgery Detection in Hyperspectral Document Images ............................. 496  
Zhipei Luo, Faisal Shafait and Ajmal Mian

Towards Query-by-Speech Handwritten Keyword Spotting ................................. 501  
Marçal Rusiñol, David Aldavert, Ricardo Toledo and Josep Llados

True Color Distributions of Scene Text and Background ....................................... 506  
Renwu Gao, Shoma Eguchi and Seiichi Uchida

Blind Versus Unblind Performance Evaluation of Binarization Methods .................... 511  
Amina Djema, Youcef Chibani, Abdenour Sehad and Et-Tahir Zemouri

Representation and Reconstruction of Map Regions ........................................... 516  
Samit Biwas, Sekhar Mandal and Amit Kumar Das

Fisher Vector Encoding of Micro Color Features for (Real World) Jigsaw Puzzles ... 521  
Fabian Richter, Christian Eggert and Rainer Lienhart

Recognizing Perspective Scene Text with Context Feature ..................................... 526  
Anna Zhu, Yangbo Dong and Guoyou Wang

Automatic Script Identification in the Wild ..................................................... 531  
Baoguang Shi, Cong Yao, Chengquan Zhang, Xiaowei Guo, Feiyue Huang and Xiang Bai
Influence of Text Line Segmentation in Handwritten Text Recognition ............................... 536
Verónica Romero, Joan Andreu Sánchez, Vicente Bosch, Katrien Depuydt and Jesse de Does

Effects of Clustering Algorithms on Typographic Reconstruction ....................................... 541
Elisa H. Barney Smith and Bart Lamirey

Chinese Character-level Writer Identification using Path Signature Feature, DropStroke, and
Deep CNN ......................................................................................................................... 546
Weixin Yang, Lianwen Jin and Manfei Liu

Improved Deep Convolutional Neural Network For online Handwritten Chinese Character
Recognition using Domain-Specific Knowledge .................................................................. 551
Weixin Yang, Lianwen Jin, Zecheng Xie and Ziyong Feng

OCR performance prediction using cross-OCR alignment ................................................. 556
Ahmed Ben Salah, Jean-Philippe Moreux, Nicolas Ragot and Thierry Paquet

Shape-based Word Spotting in Handwritten Document Images ......................................... 561
Angelos P. Giotis, Giorgos Sfikas, Christohoros Nikou and Basilis Gatos

Multiresolution Approach Based on Adaptive Superpixels for Administrative Documents
Segmentation into Color Layers ......................................................................................... 566
Elodie Carel, Jean-Christophe Burie, Vincent Courboulay, Jean-Marc Ogier and Vincent
Poulin d’Andecy

Text-Graphics Separation to Detect Logo and Stamp from Color Document Images: A Spectral
Approach ............................................................................................................................... 571
Amit Nandedkar, Jayanta Mukhopadhyay and Shamik Sural

A Conditional Random Field Model for Font Forgery Detection ....................................... 576
Romain Bertrand, Oriol Ramos Terrades, Jean-Marc Ogier, Petra Gomez-Krämer and Patrick
Franco

Parallel Sequence Classification using Recurrent Neural Networks and Alignment .................. 581
Federico Raue, Wonmin Byeon, Thomas Breuel and Marcus Livicki

One-shot field spotting on colored forms using subgraph isomorphism .............................. 586
Maroua Hammami, Pierre Héraux, Sébastien Adam and Vincent Poulin d’Andecy

DASyR(IR) - Document Analysis System for Systematic Reviews (for Information Retrieval) ..... 591
Florina Piroi, Aldo Lipani, Mihai Lupu and Allan Hanbury

A Comparative Study of Local Detectors and Descriptors for Mobile Document Classification..... 596
Marçal Rusiñol, Joseph Chazalon, Jean-Marc Ogier and Josep Llados

SRIF: Scale and Rotation Invariant Features for Camera-Based Document Image Retrieval .... 601
Quoc Bao Dang, Muhammad Muzzamil Luqman, Mickaël Coustaty, Cao De Tran and
Jean-Marc Ogier

Segmentation-Free Pattern Spotting in Historical Document Images .................................... 606
Sovann En, Caroline Petitjean, Stephane Nicolas and Laurent Heutte

A Complete Automatic Short Answer Assessment System With Student Identification ........... 611
Hemmaphan Suwanwiwat, Umapada Pal and Michael Blumenstein
Proceedings of 13th International Conference on Document Analysis and Recognition (ICDAR’2015)

Viral Transcript Alignment ................................................................. 711
* Gil Sadeh, Lior Wolf, Tal Hassner, Daniel Stoekl and Nachum Dershowitz *

Sparsely Sampled Binary Patterns for writer identification ............................................................ 716
* Anguelos Nicolaou, Andrew David Bagdanov, Marcus Liwicki and Dimosthenis Karatzas *

Use Case Visual Bag-of-Words Techniques for Camera Based Identity Document Classification ............ 721
* Lluís-Pere De Las Heras, Oriol Ramos Terrades, Josep Lladós, David Fernández and Cristina Cañero *

Attributed Graph Grammar for Floor Plan Analysis ................................................................. 726
* Lluís-Pere De Las Heras, Oriol Ramos Terrades and Josep Lladós *

Probabilistic Interpretation and Improvements to the HMM-Filler for Handwritten Keyword Spotting ................................................................. 731
* Joan Puigcerver, Alejandro Héctor Toselli and Enrique Vidal *

Context-Aware Lattice based Filler approach for Key Word Spotting in Handwritten Documents ....... 736
* Alejandro Héctor Toselli, Joan Puigcerver and Enrique Vidal *

High Performance Query-by-Example Keyword Spotting Using Query-by-String Techniques ............. 741
* Enrique Vidal, Alejandro Héctor Toselli and Joan Puigcerver *

Efficient Scene Text Localization and Recognition with Local Character Refinement ....................... 746
* Lukáš Neumann and Jiří Matas *

Multi-stage HMM based Arabic text recognition with rescoring .................................................... 751
* Irfan Ahmad and Gernot Fink *

Label Transition and Selection Pruning and Automatic Decoding Parameter Optimization for 
Time-Synchronous Viterbi Decoding ......................................................................................... 756
* Yasuhisa Fujii, Dmitriy Genzel, Ashok C. Popat and Remco Teunen *

Content-based Comic Retrieval Using Multilayer Graph Representation and Frequent Graph Mining ................................................................. 761
* Thanh Nam Le, Muhammad Muazzam Luqman, Jean-Christophe Burie and Jean-Marc Ogier *

Recognition of Historical Greek Polytonic Scripts Using LSTM Networks ..................................... 766
* Fotini Simistira, Adnan Ul Hassan, Vassilis Papavassiliou, Basilis Gatos, Vassilis Katsouros and Marcus Liwicki *

Document Image OCR Accuracy Prediction via Latent Dirichlet Allocation .................................... 771
* Xujun Peng, Huaigu Cao and Prem Natarajan *

Text Zone Classification using Unsupervised Feature Learning ..................................................... 776
* Nibal Nayef and Jean-Marc Ogier *

Handwritten Word Spotting by Inexact Matching of Grapheme Graphs ........................................ 781
* Pau Riba Fiérez, Josep Llados and Alicia Fornés *

Localized Document Image Change Detection ............................................................................. 786
* Rajiv Jain and David Doermann *

Overlapped-Triangle Analysis with Hierarchical Ranking of Dominance ...................................... 791
* Xiaqing Lu, Lu Liu, Zhi Tang, Haibin Ling and Jingwei Qu
Automated Analysis of Line Plots in Documents ................................................................. 796
Rathin Radhakrishnan Nair, Nishant Sankaran, Ifeoma Nwogu and Venu Govindaraju

Chart Classification By Combining Deep Convolutional Networks and Deep Belief Networks ...... 801
Liu Xiao, Tang Binbin, Wang Zhenyang, Xu Xianghua, Pu Shiliang, Tao Dapeng and Song Mingli

A Character Degradation Model for Color Document Images .............................................. 806
Do Thi Luyen, Elodie Carel, Jean-Marc Ogier and Jean-Christophe Burie

Multilingual Signature Verification by Combined Segmentation Verification ......................... 811
Wataru Ohyama, Yuuki Ogi, Tetsushi Wakabayashi and Fumitaka Kimura

Tackling Pattern Recognition by Vector Space Embedding .................................................... 816
Brian Iwana, Seiichi Uchida, Kaspar Riesen and Volkmar Frinken

MRF Based Text Binarization in Complex Images using Stroke Feature ................................. 821
Yanna Wang, Cunzhao Shi, Baihua Xiao and Chunheng Wang

Simplifying The Reading of Historical Manuscripts ............................................................... 826
Abedelkadir Asi, Rafi Cohen, Klara Kedem and Jihad El-Sana

Optical Modelling and Language Modelling Trade-off for Handwritten Text Recognition .......... 831
Mauricio Villegas, Joan Andreu Sanchez and Enrique Vidal

ALTID : Arabic/Latin Text Images Database for recognition research .................................... 836
Imen Chtourou, Ahmed Cheikh Rouhou, Faten Kallel and Slim Kanoun

Writer Adaptive Feature Extraction Based on Convolutional Neural Networks For Online
Handwritten Chinese Character Recognition ........................................................................... 841
Jun Du, Jian-Fang Zhai, Jin-Shui Hu, Bo Zhu, Si Wei and Lirong Dai

High Performance Offline Handwritten Chinese Character Recognition Using GoogLeNet and
Directional Feature Maps ......................................................................................................... 846
Zhuoyao Zhong, Lianwen Jin and Zecheng Xie

No threshold, no parameter. A prelude ....................................................................................... 851
Sébastien Eskenazi, Petra Gomez-Krämer and Jean-Marc Ogier

Comic Frame Extraction via Line Segments Combination ...................................................... 856
Yongtao Wang, Yafeng Zhou and Zhi Tang

Mixed handwritten and printed digit recognition in Sudoku with Convolutional Deep Belief
Network .................................................................................................................................... 861
Baptiste Wicht and Jean Hennebert

Date Field Extraction from Handwritten Documents Using HMMs ......................................... 866
Ranju Mandal, Partha Roy, Umapada Pal and Michael Blumenstein

Joint Denoising and Magnification of Noisy Low-Resolution Textual Images ....................... 871
Rim Walha, Fadoua Drira, Franck Lebourgeois, Adel M. Alimi and Christophe Garcia

Efficient Word Image Retrieval using Fast DTW Distance ..................................................... 876
Nagendar G and C.V Jawahar

Query by string word spotting based on character bi-gram indexing ..................................... 881
Suman Ghosh and Ernest Valveny
Automatic Discrimination of Text and Non-Text Natural Images ......................................................... 886
Chengquan Zhang, Cong Yao, Baoguang Shi and Xiang Bai

Confidence Measures for Seamless Skew and Orientation Detection in Document Images ........ 891
Iuliu Konya, Stefan Eickeler and Christian Brandt

Arabic Ligatures: Analysis and Application in Text Recognition ....................................................... 896
Yousef Elarian, Irfan Ahmad, Sameh Awaida, Wasfi Al-Khatib and Abdelmalek Zidouri

A clump splitting based method to localize speech balloons in comics ........................................... 901
Xicheng Liu, Yongtao Wang and Zhi Tang

Writer Identification Using VLAD Encoded Contour-Zernike Moments ............................................. 906
Vincent Christlein, David Bernecker and Elli Angelopoulou

Deep BLSTM Neural Networks for Unconstrained Continuous Handwritten Text Recognition .... 911
Volkmar Frinken and Seiichi Uchida

Robust Seed-Based Stroke Width Transform for Text Detection in Natural Images ............................ 916
Feng Su and Hailiang Xu

Interactive Content-Based Document Retrieval Using Fuzzy Attributed Relational Graph Matching .............................................................. 921
Ramzi Chaieb, Karim Kalti and Najoua Essoukri Ben Amara

A New Wavelet-Laplacian Method for Arbitrarily-Oriented Character Segmentation in Video Text Lines ........................................................................ 926
Guozhu Liang, Palaiahnakote Shivakumara, Tong Lu and Chew Lim Tan

The ENP Image and Ground Truth Dataset of Historical Newspapers ............................................. 931
Christian Clausner, Stefan Pletschacher, Christos Papadopoulos and Apostolos Antonacopoulos

A hypothesize-and-verify framework for text recognition using Deep Recurrent Neural Network ... 936
Anupama Ray, Sai Rajeswar and Santanu Chaudhury

Arabic handwritten texts clusterization based on feature relation graph (FRG) .................................... 941
Vladislav Pavlov and Dmitry Shalymov

A Segmentation-Free Approach for Printed Devanagari Script Recognition ........................................ 946
Tushar Karayil, Adnan Ul-Hasan and Thomas Breuel

Multi-Lingual Text Recognition from Video Frames ................................................................. 951
Nabin Sharma, Ranju Mandal, Rabi Sharma, Partha Partim Roy, Umapada Pal and Michael Blumenstein

Crossing the lines: making optimal use of context in line-based Handwritten Text Recognition .......... 956
Jafar Tanha, Jesse de Does, Katrien Depuydt and Joan Andreu Sánchez

Color Structure Recovering in Strong Specular Text Regions ............................................................. 961
Tam Nguyen

A Sigma-Lognormal Model For Character Level handwritten CAPTCHA Generation ............. 966
Chetan Ramaiah, Réjean Plamondon and Venu Govindaraju

Noise Characterization in Ancient Document Images Based on DCT Coefficient Distribution ....... 971
Fitri Arnia, Fardian Fardian, Sayed Muchallil and Khairul Munadi
Can RNNs Reliably Separate Script and Language at Word and Line Level? ................. 976
Ajeet Kumar Singh and Jawahar C V
Visual Appearance based Document Classification Methods: Performance Evaluation and Benchmarking ................................................................. 981
Syed Saqib Bukhari and Andreas Dengel
Adapting Off-the-Shelf CNNs for Word Spotting & Recognition ......................... 986
Arjun Sharma and Pramod Sankar K.
Evaluation of Deep Convolutional Nets for Document Image Classification and Retrieval .......... 991
Adam Harley, Alex Ufkes and Konstantinos Derpanis
A Dataset for Arabic Text detection, tracking and recognition in news VideosAcTiV .......... 996
Oussama Zayene, Jean Hennebert, Sameh Masmoudi Touj, Rolf Ingold and Najoua Essoukri Benamara
Curriculum Learning for Printed Text Line Recognition of Ligature-based Scripts .......... 1001
Adnan Ul-Hasan, Faisal Shafait and Marcus Liwicki
Gradient-Domain Degradations for Improving Historical Documents Images Layout Analysis .... 1006
Mathias Seuret, Kai Chen, Nicole Eichenberger, Marcus Liwicki and Rolf Ingold
Page Segmentation of Historical Document Images with Convolutional Autoencoders ........ 1011
Kai Chen, Mathias Seuret, Marcus Liwicki, Jean Hennebert and Rolf Ingold
Word of Blobs ................................................................. 1016
Jihad El-Sana and Klara Kedem
CNN Based Common Approach to Handwritten Character Recognition of Multiple Scripts .... 1021
Durjoy Sen Maitra, Ujjwal Bhattacharya and Swapan Kr. Parui
Deep Learning and Recurrent Connectionist-based Approaches for Arabic Text Recognition in Videos ................................................................. 1026
Sonia Yousfi, Sid-Ahmed Berrani and Christophe Garcia
Planar Markovian Approach for the Recognition of a Wide Vocabulary of Arabic Decomposable Words ................................................................. 1031
Imen Ben Cheikh and Imen Allagui
Arabic Characters Recognition in Natural Scenes using Sparse Coding for Feature Representations. 1036
Maroua Toumi, Ikram Moalla, Adel M. Alimi and Frank Lebourgais
Unsupervised Feature Learning for Optical Character Recognition ............................ 1041
Devendra Sahu and Jawahar C. V.
A Sequence Learning Approach for Multiple Scripts Identification .......................... 1046
Adnan Ul-Hasan, Muhammad Zeshan Afzal, Faisal Shafait, Marcus Liwicki and Thomas Breuel
Trajectory Recovery and Stroke Reconstruction of Mathematical Symbols .................. 1051
Behrang Sabeghi Saroui and Volker Sorge
Unconstrained Bengali Handwriting Recognition with Recurrent Models .................... 1056
Utpal Garain, Luc Mioulet, Bidyut Baran Chaudhuri, Clement Chatelain and Thierry Paquet
Online Handwriting Recognition using Depth Sensors ............................................. 1061
Rajat Aggarwal, Sirnam Swetha, Anoop M. Namboodiri, Jayanthy Sivaswamy and C. V. Jawahar
ICDAR2015 Competition on Recognition of Documents with Complex Layouts - RDCL2015

Apostolos Antonacopoulos, Christian Clausner, Christos Papadopoulos and Stefan Pletschacher

ICDAR 2015 Competition on Robust Reading


ICDAR2015 Competition on Smartphone Document Capture and OCR (SmartDoc)

Jean-Christophe Burie, Joseph Chazalon, Mickael Coustaty, Sebastien Eskenazi, Muhammad Muzzamil Luqman, Maroua Mehri, Nibal Nayef, Jean-Marc Ogier, Sophea Prum and Marcal Rusinol

ICDAR 2015 Competition HTRtS: Handwritten Text Recognition on the tranScriptorium Dataset

Joan Andreu Sanchez, Alejandro H. Toselli, Veronica Romero and Enrique Vidal

ICDAR 2015 Text Line Detection in Historical Documents (ANDAR-TL-2015)

Michael Murdock

ICDAR2015 Competition on Keyword Spotting for Handwritten Documents

Joan Puigcerver, Alejandro Hector Toselli and Enrique Vidal

ICDAR 2015 Contest on MultiSpectral Text Extraction (MS-TEX 2015)

Rachid Hedjam, Hossein Ziaei Nafchi, Reza Farrahi Moghaddam, Margaret Kalacska and Mohamed Cheriet

ICDAR 2015 Competition on Signature Verification and Writer Identification for On- and Off-line Skilled Forgeries (SigWIcomp2015)

Muhammad Imran Malik

ICDAR2015 Competition on Multi-script Writer Identification and Gender Classification using QUWI Database

Chawki Djeddi, Somaya Al-Maadeed, Abdeljalil Gattal, Imran Siddiqi, Labiba Souici-Meslati and Haikal El Abed

ICDAR 2015 Competition on Video Script Identification (CVSI 2015)

Nabin Sharma, Ranju Mandal, Rabi Sharma, Umapada Pal and Michael Blumenstein

ICDAR2015 Competition on Text Image Super-Resolution

Clément Peyrard, Moez Baccouche, Franck Mamalet and Christophe Garcia

Workshops Papers

AMIGO – Automatic Indexing of Lecture Footage

Markus Eberts, Adrian Ulges and Ulrich Schwanecke

Camera-based Document Image Retrieval System using Local Features - comparing SRIF with LLAH, SIFT, SURF and ORB

Q.B. Dang, V.P. Le, M.M. Luqman, M. Coustaty, C.D. Tran and J-M. Ogier

Improving Document Matching Performance by Local Descriptor Filtering

Joseph Chazalon, Marc al Rusinol and Jean-Marc Ogier

ALIF: A Dataset for Arabic Embedded Text Recognition in TV Broadcast

Sonia Yousfi, Sid-Ahmed Berrani and Christophe Garcia

xiv
Proceedings of 13th International Conference on Document Analysis and Recognition (ICDAR’2015)

Rectification of Camera Captured Document Images for Camera-Based OCR Technology ........ 1226
Mohamed Fawzi, Mohsen A. Rashwan, Hany Ahmed, shaimaa Samir, Sherif M. Abdou, Hassanin M. Al-Barhamoshy and Kamal M. Jambi

SmartDoc-QA: A Dataset for Quality Assessment of Smartphone Captured Document Images - Single and Multiple Distortions ................................................................. 1231
Nibal Nayef, Muhammad Muzzamil Luqman, Sophea Prum, Sebastien Eskenazi, Joseph Chazalon and Jean-Marc Ogier

Efficient indexing for Query By String text retrieval ......................................................... 1236
Suman K. Ghosh Lluis Gomez, Dimosthenis Karatzas and Ernest Valveny

Multi-script Iterative Steerable Directional Filtering For Handwritten Text Line Extraction ........ 1241
Wassim Swaileh, Kamel Ait Mohand and Thierry Paquet

Recognizable Units in Pashto Language for OCR ............................................................. 1246
Riaz Ahmad, Muhammad Zeshan Afzal, Sheikh Faisal Rashid, Marcus Liwicki, Andreas Dengel and Thomas Breuel

Script Independent Online Handwriting Recognition ......................................................... 1251
Oendrila Samanta, Anandarup Roy, Ujjwal Bhattacharya and Swapan K. Parui

OCR for Bilingual documents using Language Modeling .................................................. 1256
Anupama Ray, Sai Rajeswar and Santanu Chaudhury

Document Indexing Framework for Retrieval of Degraded Document Images ..................... 1261
Ritu Garg, Ehtesham Hassan and Santanu Chaudhury

Quantitative Evaluation of Features for Forensic Handwriting Examination ........................ 1266
Angelo Marcelli, Antonio Parziale and Claudio De Stefano

Handedness Detection of Online Handwriting based on Horizontal Strokes ........................ 1272
Erika Griechisch and Erika Bencsik

Behaviour of Dynamic and Static Feature Dependences in Constrained Signatures ............... 1278
Giuseppe Pirlo, Moises Diaz, Miguel A. Ferrer, Donato Impedovo and Fabrizio Rizzi