

*Proceedings of the*

# 8<sup>th</sup> IAPR International Workshop on Document Analysis Systems

September 16-19, 2008

Nara, Japan



Los Alamitos, California

Washington • Tokyo



# TABLE OF CONTENTS

<b>Extraction of Text Objects in Video Documents: Recent Progress</b> .....	1
<i>Jing Zhang, Rangachar Kasturi</i>	
<b>A Hilbert Warping Algorithm for Recognizing Characters from Moving Camera</b> .....	14
<i>Hiroyuki Ishida, Ichiro Ide, Hiroshi Murase, Tomokazu Takahashi</i>	
<b>Writer Verification of Arabic Handwriting</b> .....	21
<i>Sargur N. Srihari, Gregory R. Ball</i>	
<b>A Robust System to Detect and Localize Texts in Natural Scene Images</b> .....	28
<i>Yi-Feng Pan, Xinwen Hou, Cheng-Lin Liu</i>	
<b>An Image Based Watermark String Detection System for Document Security Checking</b> .....	36
<i>Jun Sun, Satashi Naoi, Yusaka Fujii, Hiroaki Takebe, Katsuhito Fujimoto</i>	
<b>Feature Extraction for Document Image Segmentation by pLSA Model</b> .....	44
<i>Takuma Yamaguchi, Minoru Maruyama</i>	
<b>Grouping Text Lines in Online Handwritten Japanese Documents by Combining Temporal and Spatial Information</b> .....	52
<i>Xiang-Dong Zhou, Da-Han Wang, Cheng-Lin Liu</i>	
<b>Accurate Alignment of Double-Sided Manuscripts for Bleed-Through Removal</b> .....	60
<i>Jie Wang, Michael S. Brown, Chew Lim Tan</i>	
<b>Super-Resolution of Text Images Using Edge-Directed Tangent Field</b> .....	67
<i>Jyotirmoy Banerjee, C.V. Jawahar</i>	
<b>Attention-Based Document Classifier Learning</b> .....	75
<i>Georg Buscher, Andreas Dengel</i>	
<b>Categorization of On-Line Handwritten Documents</b> .....	83
<i>Sebastián Peña Saldarriaga, Emmanuel Morin, Christian Viard-Gaudin</i>	
<b>Combining Multiple Methods for Book Indexing</b> .....	91
<i>Hervé Déjean, Jean-Luc Meunier</i>	
<b>Automated OCR Ground Truth Generation</b> .....	99
<i>Joost van Beusekom, Faisal Shafait, Thomas M. Breuel</i>	
<b>Digital Renaissance—Making Archives, Sharing Wisdoms and Creating Values</b> .....	106
<i>Toshiro Kamiuchi</i>	
<b>State: A Multimodal Assisted Text-Transcription System for Ancient Documents</b> .....	118
<i>Albert Gordo, David Llorens, Andrés Marzal, Federico Prat, Juan Miguel Vilar</i>	
<b>Authorship Identification of Ukiyoe by Using Rakkan Image</b> .....	126
<i>Shun Hirose, Mitsu Yoshimura, Kozaburo Hachimura, Ryo Akama</i>	
<b>Writer-Dependent Recognition of Handwritten Whiteboard Notes in Smart Meeting Room Environments</b> .....	134
<i>Marcus Liwicki, Andreas Schlapbach, Horst Bunke</i>	
<b>Shape Code Based Lexicon Reduction for Offline Handwritten Word Recognition</b> .....	141
<i>Roman Bertolami, Christoph Gutmann, Horst Bunke, A. Lawrence Spitz</i>	
<b>A Document Analysis System for Supporting Electronic Voting Research</b> .....	147
<i>Daniel Lopresti, George Nagy, Elisa Barney Smith</i>	

<b>An End-to-End Administrative Document Analysis System</b> .....	155
<i>Hatem Hamza, Yolande Belaïd, Abdel Belaïd, Bidyut B. Chaudhuri</i>	
<b>Re-targetable OCR with Intelligent Character Segmentation</b> .....	163
<i>Mudit Agrawal, David Doermann</i>	
<b>Symbol Descriptor Based on Shape Context and Vector Model of Information Retrieval</b> .....	171
<i>T-O Nguyen, S. Tabbone, O. Ramos Terrades</i>	
<b>Skew Estimation by Instances</b> .....	178
<i>Seiichi Uchida, Megumi Sakai, Masakazu Iwamura, Shinichiro Omachi, Koichi Kise</i>	
<b>A Two-Step Dewarping of Camera Document Images</b> .....	186
<i>Nikolaos Stamatoopoulos, Basilis Gatos, Ioannis Pratikakis, Stavros J. Perantonis</i>	
<b>An Objective Evaluation Methodology for Document Image Binarization Techniques</b> .....	194
<i>K. Ntirogiannis, B. Gatos, I. Pratikakis</i>	
<b>Contrast Enhancement in Multispectral Images by Emphasizing Text Regions</b> .....	202
<i>Martin Lettner, Florian Kleber, Robert Sablatnig, Heinz Miklas</i>	
<b>Affine Invariant Recognition of Characters by Progressive Pruning</b> .....	210
<i>Akira Horimatsu, Ryo Niwa, Masakazu Iwamura, Koichi Kise, Seiichi Uchida, Shinichiro Omachi</i>	
<b>Detecting Gradients in Text Images Using the Hough Transform</b> .....	218
<i>Dimosthenis Karatzas</i>	
<b>Multi-Font Rotated Character Recognition Using Periodicity</b> .....	226
<i>Hiroyuki Hase, Kohei Tanabe, Thi Hong Ha Tran, Shogo Tokai</i>	
<b>Difference of Boxes Filters Revisited: Shadow Suppression and Efficient Character Segmentation</b> .....	234
<i>Erik Rodner, Herbert Süsse, Wolfgang Ortmann, Joachim Denzler</i>	
<b>Segmentation of Curled Textlines Using Active Contours</b> .....	241
<i>Syed Saqib Bukhari, Faisal Shafait, Thomas M. Breuel</i>	
<b>Kanji Character Detection from Complex Real Scene Images based on Character Properties</b> .....	249
<i>Lianli Xu, Hiroto Nagayoshi, Hiroshi Sako</i>	
<b>A Hybrid System for Text Detection in Video Frames</b> .....	257
<i>Marios Anthonopoulos, Basilis Gatos, Ioannis Pratikakis</i>	
<b>A Study for High Performance Character Extraction from Color Scene Images</b> .....	264
<i>Keiichiro Shirai, Masanori Wakabayashi, Masayuki Okamoto, Hiroaki Yamamoto</i>	
<b>Word Extraction Method by Generating Multiple Character Hypotheses</b> .....	270
<i>Hiroaki Takebe, Katsuhito Fujimoto</i>	
<b>An Efficient Edge Based Technique for Text Detection in Video Frames</b> .....	278
<i>Palaiahnakote Shivakumara, Weihua Huang, Chew Lim Tan</i>	
<b>Multi-Oriented English Text Line Extraction Using Background and Foreground Information</b> .....	286
<i>Partha Pratim Roy, Umapada Pal, Josep Lladós, Fumitaka Kimura</i>	
<b>Text String Extraction from Scene Image Based on Edge Feature and Morphology</b> .....	294
<i>Yuming Wang, Naoki Tanaka</i>	
<b>Accuracy Improvement and Objective Evaluation of Annotation Extraction from Printed Documents</b> .....	300
<i>Tomohiro Nakai, Kazumasa Iwata, Koichi Kise</i>	

<b>Multi-Oriented Text Line Extraction from Handwritten Arabic Documents</b> .....	308
<i>Nazih Ouwayed, Abdel Belaid</i>	
<b>Writer Identification in Old Handwritten Music Scores</b> .....	316
<i>Alicia Fornés, Josep Lladós, Gemma Sánchez, Horst Bunke</i>	
<b>An Expansion Method for Off-line Hand-printed Character Recognition using On-line Character Writing Features</b> .....	323
<i>Hiromitsu Nishimura, Kazuhisa Yanaka</i>	
<b>A Model-Based Field Frame Detection for Handwritten Filled-in Forms</b> .....	334
<i>Juan-Carlos Perez-Cortes, Luis Andreu, Joaquim Arlandis</i>	
<b>Lexicon Reduction in Handwriting Recognition Using Topic Categorization</b> .....	341
<i>Faisal Farooq, Gaurav Chandalia, Venu Govindaraju</i>	
<b>Truthing for Pixel-Accurate Segmentation</b> .....	348
<i>Michael A. Moll, Henry S. Baird, Chang An</i>	
<b>On the Reading of Tables of Contents</b> .....	355
<i>Prateek Sarkar, Eric Saund</i>	
<b>Unsupervised Decomposition of Color Document Images by Projecting Colors to a Spherical Surface</b> .....	363
<i>Yuan He, Jun Sun, Satoshi Naoi, Yusaku Fujii, Katsuhito Fujimoto</i>	
<b>Fast and Accurate Skew Estimation Based on Distance Transform</b> .....	371
<i>Itay Bar-Yosef, Nate Hagbi, Klara Kedem, Itshak Dinstein</i>	
<b>Physical Layout Segmentation of Mail Application Dedicated to Automatic Postal Sorting System</b> .....	377
<i>Djamel Gaceb, Véronique Eglin, Frank LeBourgeois, Hubert Emptoz</i>	
<b>Structural Mixtures for Statistical Layout Analysis</b> .....	384
<i>Faisal Shafait, Joost van Beusekom, Daniel Keysers, Thomas M. Breuel</i>	
<b>A Study on Document Structure Recognition of Discharge Summaries for Analogous Case Search System</b> .....	392
<i>Hiroharu Kawanaka, Yujiro Shiroyama, Shinji Tsuruoka, Tsuyoshi Shinogi, Koji Yamamoto</i>	
<b>A Fast Preprocessing Method for Table Boundary Detection: Narrowing Down the Sparse Lines using Solely Coordinate Information</b> .....	400
<i>Ying Liu, Prasenjit Mitra, C. Lee Giles</i>	
<b>Pre-Printed and Hand-Filled Table-Form Analysis Aiming Cell Extraction</b> .....	408
<i>Rafaela Dandolini Felipe, Luiz Antonio Pereira Neves</i>	
<b>Efficient Binarization of Historical and Degraded Document Images</b> .....	413
<i>B. Gatos, I. Pratikakis, S.J. Perantonis</i>	
<b>A Graphics Image Processing System</b> .....	421
<i>Linlin Li, Chew Lim Tam</i>	
<b>A Proposal of Evaluation Method for Balance of White Space in Calligraphy by Using Horizon View Camera</b> .....	429
<i>Kensuke Tobitani, Nobuyasu Okabe, Kazuhiko Yamamoto, Kunihito Kato</i>	
<b>Anisotropic Total Variation Method for Text Image Super-Resolution</b> .....	436
<i>Battulga Bayarsaikhan, Youghee Kwon, Jin Hyung Kim</i>	
<b>CCD: Connected Component Descriptor for Robust Mosaicing of Camera-Captured Document Images</b> .....	443
<i>T. Kasar, A.G. Ramakrishnan</i>	

<b>Word and Symbol Spotting Using Spatial Organization of Local Descriptors</b> .....	450
<i>Marçal Rusiñol, Josep Lladós</i>	
<b>Performance Evaluation of Symbol Recognition and Spotting Systems: An Overview</b> .....	458
<i>Mathieu Delalandre, Ernest Valveny, Josep Lladós</i>	
<b>Object Extraction from Colour Cadastral Maps</b> .....	467
<i>Romain Raveaux, Jean-Christophe Burie, Jean-Marc Ogier</i>	
<b>HistoSketch: A Semi-Automatic Annotation Tool for Archival Documents</b> .....	476
<i>Joan Mas, Jose A. Rodriguez, Dimosthenis Karatzas, Gemma Sanchez, Josep Lladós</i>	
<b>A Complete Optical Character Recognition Methodology for Historical Documents</b> .....	484
<i>G. Vamvakas, B. Gatos, N. Stamatopoulos, S.J. Perantonis</i>	
<b>Document Image Retrieval to Support Reading Mekkans</b> .....	492
<i>Akihito Kitadai, Jun Takakura, Masatoshi Ishikawa, Masaki Nakagawa, Hajime Baba, Akihiro Watanabe</i>	
<b>Keyword Matching in Historical Machine-Printed Documents Using Synthetic Data, Word Portions and Dynamic Time Warping</b> .....	498
<i>Thomas Konidakis, B. Gatos, S.J. Perantonis, A. Kesidis</i>	
<b>A Large-Scale Analysis of Mathematical Expressions for an Accurate Understanding of Their Structure</b> .....	505
<i>Walaa Aly, Seiichi Uchida, Masakazu Suzuki</i>	
<b>An Empirical Measure on the Set of Symbols Occurring in Engineering Mathematics Texts</b> .....	513
<i>Stephen M. Watt</i>	
<b>Named Entity Recognition by Neural Sliding Window</b> .....	521
<i>Ignazio Gallo, Elisabetta Binaghi, Moreno Carullo, Nicola Lamberti</i>	
<b>Exploring Evolutionary Technical Trends from Academic Research Papers</b> .....	528
<i>Teng-Kai Fan, Chia-Hui Chang</i>	
<b>PaperDiff: A Script Independent Automatic Method for Finding the Text Differences Between Two Document Images</b> .....	536
<i>Sitaram Ramachandrula, Gopal Datt Joshi, Noushath. S, Pulkit Parikh, Vishal Gupta</i>	
<b>The HCI Paradigm of HyperPrinting</b> .....	542
<i>Thomas Kieninger, Andreas Dengel</i>	
<b>MathBrush: A System for Doing Math on Pen-Based Devices</b> .....	550
<i>George Labahn, Edward Lank, Scott MacLean, Mirette Marzouk, David Tausky</i>	
<b>End-to-End Trainable Thai OCR System Using Hidden Markov Models</b> .....	558
<i>Kriste Krstovski, Ehry MacRostie, Rohit Prasad, Premkumar Natarajan</i>	
<b>Comprehensive Global Typography Extraction System for Electronic Book Documents</b> .....	566
<i>Liangcai Gao, Zhi Tang, Xiaofan Lin, Ruiheng Qiu</i>	
<b>Digital Ink to Form Alignment for Electronic Clipboard Devices</b> .....	573
<i>Jagannadan Varadarajan, Sriganesh Madhvanath</i>	
<b>Towards Whole-Book Recognition</b> .....	580
<i>Pingping Xiu, Henry S. Baird</i>	
<b>Handling of Surface Modifications for Robust Image Based Mail Piece Comparison</b> .....	588
<i>Katja Worm, Beate Meffert</i>	
<b>Dolores: An Interactive and Class-Free Approach for Document Logical Restructuring</b> .....	595
<i>Jean-Luc Bloechle, Catherine Pugin, Rolf Ingold</i>	

<b>Hierarchical Indexing and Semantic Annotation for Document Representations in Community-Web Virtual Libraries</b> .....	604
<i>Meriem Fellah, Anja Habacha Chaibi, Mohamed Ben Ahmed</i>	
<b>The Convergence of Iterated Classification</b> .....	612
<i>Chang An, Henry S. Baird</i>	
<b>A Comparison of Clustering Methods for Word Image Indexing</b> .....	620
<i>Simone Marinai, Emanuele Marino, Giovanni Soda</i>	
<b>New Oversampling Approaches Based on Polynomial Fitting for Imbalanced Data Sets</b> .....	626
<i>Sami Gazzah, Najoua Essoukri Ben Amara</i>	
<b>Author Index</b>	