

Table of Contents

OCR Features and Systems

Relating Statistical Image Differences and Degradation Features	1
<i>Elisa Barney Smith, Xiaohui Qiu</i>	
Script Identification in Printed Bilingual Documents	13
<i>D. Dhanya, A.G. Ramakrishnan</i>	
Optimal Feature Extraction for Bilingual OCR	25
<i>D. Dhanya, A.G. Ramakrishnan</i>	
Machine Recognition of Printed Kannada Text	37
<i>B. Vijay Kumar, A.G. Ramakrishnan</i>	
An Integrated System for the Analysis and the Recognition of Characters in Ancient Documents	49
<i>Stefano Vezzosi, Luigi Bedini, Anna Tonazzini</i>	
A Complete Tamil Optical Character Recognition System	53
<i>K.G. Aparna, A.G. Ramakrishnan</i>	
Distinguishing between Handwritten and Machine Printed Text in Bank Cheque Images	58
<i>José Eduardo Bastos Dos Santos, Bernard Dubuisson, Flávio Bortolozzi</i>	
Multi-expert Seal Imprint Verification System for Bankcheck Processing	62
<i>Katsuhiko Ueda, Ken'ichi Matsuo</i>	
Automatic Reading of Traffic Tickets	66
<i>Nabeel Murshed</i>	

Handwriting Recognition

A Stochastic Model Combining Discrete Symbols and Continuous Attributes and Its Application to Handwriting Recognition	70
<i>Hanhong Xue, Venu Govindaraju</i>	
Top-Down Likelihood Word Image Generation Model for Holistic Word Recognition	82
<i>Eiki Ishidera, Simon M. Lucas, Andrew C. Downton</i>	

The Segmentation and Identification of Handwriting in Noisy Document Images	95
<i>Yefeng Zheng, Huiping Li, David Doermann</i>	

The Impact of Large Training Sets on the Recognition Rate of Offline Japanese Kanji Character Classifiers	106
<i>Ondrej Velek, Masaki Nakagawa</i>	

Automatic Completion of Korean Words for Open Vocabulary Pen Interface	111
<i>Sungho Ryu, Jin-Hyung Kim</i>	

Using Stroke-Number-Characteristics for Improving Efficiency of Combined Online and Offline Japanese Character Classifiers	115
<i>Ondrej Velek, Masaki Nakagawa</i>	

Closing Gaps of Discontinuous Lines: A New Criterion for Choosing the Best Prolongation	119
<i>Eugene Bodansky, Alexander Gribov</i>	

Classifiers and Learning

Classifier Adaptation with Non-representative Training Data	123
<i>Sriharsha Veeramachaneni, George Nagy</i>	

A Learning Pseudo Bayes Discriminant Method Based on Difference Distribution of Feature Vectors	134
<i>Hiroaki Takebe, Koji Kurokawa, Yutaka Katsuyama, Satoshi Naoi</i>	

Increasing the Number of Classifiers in Multi-classifier Systems: A Complementarity-Based Analysis	145
<i>L. Bovino, G. Dimauro, S. Impedovo, G. Pirlo, A. Salzo</i>	

Discovering Rules for Dynamic Configuration of Multi-classifier Systems	157
<i>G. Dimauro, S. Impedovo, M.G. Lucchese, G. Pirlo, A. Salzo</i>	

Multiple Classifier Combination for Character Recognition: Revisiting the Majority Voting System and Its Variations	167
<i>A.F.R. Rahman, H. Alam, M.C. Fairhurst</i>	

Layout Analysis

Correcting for Variable Skew	179
<i>A. Lawrence Spitz</i>	

Two Geometric Algorithms for Layout Analysis	188
<i>Thomas M. Breuel</i>	

Text/Graphics Separation Revisited	200
<i>Karl Tombre, Salvatore Tabbone, Loïc Pélissier, Bart Lamiroy, Philippe Dosch</i>	
A Study on the Document Zone Content Classification Problem	212
<i>Yalin Wang, Ihsin T. Phillips, Robert M. Haralick</i>	
Logical Labeling of Document Images Using Layout Graph Matching with Adaptive Learning	224
<i>Jian Liang, David Doermann</i>	
A Ground-Truthing Tool for Layout Analysis Performance Evaluation	236
<i>A. Antonacopoulos, H. Meng</i>	
Simple Layout Segmentation of Gray-Scale Document Images	245
<i>Aimamorn Suvichakorn, Sarin Watcharabusaracum, Wasin Sinthupinyo</i>	

Tables and Forms

Detecting Tables in HTML Documents	249
<i>Yalin Wang, Jianying Hu</i>	
Document-Form Identification Using Constellation Matching of Keywords Abstracted by Character Recognition	261
<i>Hiroshi Sako, Naohiro Furukawa, Masakazu Fujio, Shigeru Watanabe</i>	
Table Detection via Probability Optimization	272
<i>Yalin Wang, Ihsin T. Phillips, Robert M. Haralick</i>	
Complex Table Form Analysis Using Graph Grammar	283
<i>Akira Amano, Naoki Asada</i>	
Detection Approaches for Table Semantics in Text	287
<i>Saleh Alrashed, W.A. Gray</i>	
A Theoretical Foundation and a Method for Document Table Structure Extraction and Decomposition	291
<i>Howard Wasserman, Keitaro Yukawa, Bon Sy, Kui-Lam Kwok, Ihsin T. Phillips</i>	

Text Extraction

Fuzzy Segmentation of Characters in Web Images Based on Human Colour Perception	295
<i>A. Antonacopoulos, D. Karatzas</i>	
Word and Sentence Extraction Using Irregular Pyramid	307
<i>Poh Kok Loo, Chew Lim Tan</i>	

Word Searching in Document Images Using Word Portion Matching	319
<i>Yue Lu, Chew Lim Tan</i>	

Scene Text Extraction in Complex Images	329
<i>Hye-Ran Byun, Myung-Cheol Roh, Kil-Cheon Kim, Yeong-Woo Choi, Seong-Whan Lee</i>	

Text Extraction in Digital News Video Using Morphology	341
<i>Hyeran Byun, Inyoung Jang, Yeongwoo Choi</i>	

Indexing and Retrieval

Retrieval by Layout Similarity of Documents Represented with MXY Trees	353
<i>Francesca Cesarini, Simone Marinai, Giovanni Soda</i>	

Automatic Indexing of Newspaper Microfilm Images	365
<i>Qing Hong Liu, Chew Lim Tan</i>	

Improving Document Retrieval by Automatic Query Expansion Using Collaborative Learning of Term-Based Concepts	376
<i>Stefan Klink, Armin Hust, Markus Junker, Andreas Dengel</i>	

Spotting Where to Read on Pages – Retrieval of Relevant Parts from Page Images	388
<i>Koichi Kise, Masaaki Tsujino, Keinosuke Matsumoto</i>	

Mining Documents for Complex Semantic Relations by the Use of Context Classification	400
<i>Andreas Schmidt, Markus Junker</i>	

Hairetes: A Search Engine for OCR Documents	412
<i>Kazem Taghva, Jeffrey Coombs</i>	

Text Verification in an Automated System for the Extraction of Bibliographic Data	423
<i>George R. Thoma, Glenn Ford, Daniel Le, Zhirong Li</i>	

Document Engineering

<i>smartFIX</i> : A Requirements-Driven System for Document Analysis and Understanding	433
<i>Andreas R. Dengel, Bertin Klein</i>	

Multi-scale Document Description Using Rectangular Granulometries	445
<i>Andrew D. Bagdanov, Marcel Worring</i>	

Machine Learning of Generalized Document Templates for Data Extraction	457
<i>Janusz Wnek</i>	

Configuration REcognition Model for Complex Reverse Engineering Methods: 2(<i>CREM</i>)	469
<i>Karim Hadjar, Oliver Hitz, Lyse Robadey, Rolf Ingold</i>	
Electronic Document Publishing Using DjVu	480
<i>Artem Mikheev, Luc Vincent, Mike Hawrylycz, Léon Bottou</i>	
DAN: An Automatic Segmentation and Classification Engine for Paper Documents	491
<i>L. Cinque, S. Levialdi, A. Malizia, F. De Rosa</i>	
Document Reverse Engineering: From Paper to XML	503
<i>Kyong-Ho Lee, Yoon-Chul Choy, Sung-Bae Cho, Xiao Tang, Victor McCrary</i>	
New Applications	
Human Interactive Proofs and Document Image Analysis	507
<i>Henry S. Baird, Kris Popat</i>	
Data GroundTruth, Complexity, and Evaluation Measures for Color Document Analysis.	519
<i>Leon Todoran, Marcel Worring, Arnold W.M. Smeulders</i>	
Exploiting WWW Resources in Experimental Document Analysis Research	532
<i>Daniel Lopresti</i>	
An Automated Tachograph Chart Analysis System	544
<i>A. Antonacopoulos, D.P. Kennedy</i>	
A Multimodal System for Accessing Driving Directions	556
<i>Ashutosh Morde, Ramanujan S. Kashi, Michael K. Brown, Deborah Grove, James L. Flanagan</i>	
Author Index	569

Document Analysis Systems V

5th International Workshop, DAS 2002, Princeton, NJ,

USA, August 19-21, 2002. Proceedings

Lopresti, D.; Hu, J.; Kashi, R. (Eds.)

2002, XIV, 574 p., Softcover

ISBN: 978-3-540-44068-0