

Table of Contents

Digital Libraries

| | |
|----------------------------------------------------------------------------------------|----|
| Document Analysis Systems for Digital Libraries: Challenges and Opportunities | 1 |
| <i>Henry S. Baird, Venugopal Govindaraju, and Daniel P. Lopresti</i> | |
| The Trinity College Dublin 1872 Online Catalogue | 17 |
| <i>John G. Byrne</i> | |
| DL Architecture for Indic Scripts | 28 |
| <i>Suryaprakash Kompalli, Srirangaraj Sethur, and Venugopal Govindaraju</i> | |
| A Semantic-Based System for Querying Personal Digital Libraries | 39 |
| <i>Luigi Cinque, Alessio Malizia, and Roberto Navigli</i> | |
| Toward Personalized Digital Library for Providing “Information JIT” | 47 |
| <i>Hisashi Ikeda, Naohiro Furukawa, Katsumi Marukawa, and Hiromichi Fujisawa</i> | |

Historical Documents

| | |
|-------------------------------------------------------------------------------------------------------------------------|-----|
| Tilting at Windmills: Adventures in Attempting to Reconstruct <i>Don Quixote</i> | 51 |
| <i>A. Lawrence Spitz</i> | |
| A Segmentation-Free Recognition Technique to Assist Old Greek Handwritten Manuscript OCR | 63 |
| <i>Basilios Gatos, Kostas Ntzios, Ioannis Pratikakis, Sergios Petridis, T. Konidakis, and Stavros J. Perantonis</i> | |
| Automatic Metadata Retrieval from Ancient Manuscripts | 75 |
| <i>Frank Le Bourgeois and Hala Kaileh</i> | |
| A Complete Approach to the Conversion of Typewritten Historical Documents for Digital Archives | 90 |
| <i>Apostolos Antonacopoulos and Dimosthenis Karatzas</i> | |
| An Adaptive Binarization Technique for Low Quality Historical Documents | 102 |
| <i>Basilios Gatos, Ioannis Pratikakis, and Stavros J. Perantonis</i> | |
| Segmentation of Handwritten Characters for Digitalizing Korean Historical Documents | 114 |
| <i>Min Soo Kim, Kyu Tae Cho, Hee Kue Kwag, and Jin Hyung Kim</i> | |

| | |
|-----------------------------------------------------|-----|
| Self-organizing Maps and Ancient Documents | 125 |
| <i>Eddie Smigiel, Abdel Belaid, and Hatem Hamza</i> | |

| | |
|-------------------------------------------------------------|-----|
| Enriching Historical Manuscripts: The Bovary Project | 135 |
| <i>Stéphane Nicolas, Thierry Paquet, and Laurent Heutte</i> | |

Layout Analysis

| | |
|----------------------------------------------------------------------|-----|
| Word Grouping in Document Images Based on Voronoi Tessellation | 147 |
| <i>Yue Lu, Zhe Wang, and Chew Lim Tan</i> | |

| | |
|-----------------------------------------------------------------------|-----|
| Multi-component Document Image Coding Using Regions-of-Interest | 158 |
| <i>Xiao Wei Yin, Andy C. Downton, Martin Fleury, and J. He</i> | |

| | |
|-------------------------------------------------------------------------------------------------------|-----|
| Physical Layout Analysis of Complex Structured Arabic Documents Using Artificial Neural Nets | 170 |
| <i>Karim Hadjar and Rolf Ingold</i> | |

| | |
|------------------------------------------------------------------------------------------------|-----|
| An Integrated Approach for Automatic Semantic Structure Extraction in Document Images | 179 |
| <i>Margherita Berardi, Michele Lapi, and Donato Malerba</i> | |

| | |
|------------------------------------------------------------------------|-----|
| Multi-view HAC for Semi-supervised Document Image Classification | 191 |
| <i>Fabien Carmagnac, Pierre Héroux, and Éric Trupin</i> | |

| | |
|--------------------------------------------------------------------------------------|-----|
| Configurable Text Stamp Identification Tool with Application of Fuzzy Logic | 201 |
| <i>J. He and Andy C. Downton</i> | |

| | |
|------------------------------------------------------|-----|
| Layout and Content Extraction for PDF Documents..... | 213 |
| <i>Hui Chao and Jian Fan</i> | |

| | |
|---------------------------------------------------------------------|-----|
| Automatic Extraction of Filled-In Items from Bank-Check Images..... | 225 |
| <i>Katsuhiko Ueda, Hirotoshi Maegawa, and Kenichi Matsuo</i> | |

Color Documents

| | |
|------------------------------------------------------------------------------------------|-----|
| Bleed-Through Removal from Degraded Documents Using a Color Decorrelation Method..... | 229 |
| <i>Anna Tonazzini, Emanuele Salerno, Matteo Mochi, and Luigi Bedini</i> | |

| | |
|-------------------------------------------------------|-----|
| Colour Map Classification for Archive Documents | 241 |
| <i>J. He and Andy C. Downton</i> | |

| | |
|--------------------------------------------------------------------------------------------------------------------|-----|
| Serialized k -Means for Adaptative Color Image Segmentation – Application to Document Images and Others | 252 |
| <i>Yann Leydier, Frank Le Bourgeois, and Hubert Emptoz</i> | |

| | |
|--------------------------------------------------------------------------------------|-----|
| Adaptive Region Growing Color Segmentation for Text Using Irregular Pyramid | 264 |
| <i>Poh Kok Loo and Chew Lim Tan</i> | |

| | |
|--------------------------------------------------------------------------------|-----|
| Preprocessing and Segmentation of Bad Quality Machine Typed Documents | 276 |
| <i>Mariusz Szwoch and Wioleta Szwoch</i> | |

Handwritten Documents

| | |
|----------------------------------------------------------------------------------------------------------------|-----|
| Ensembles of Classifiers for Handwritten Word Recognition Specialized on Individual Handwriting Style | 286 |
| <i>Simon Günter and Horst Bunke</i> | |

| | |
|-----------------------------------------------------------------------------------------------|-----|
| Information Retrieval System for Handwritten Documents | 298 |
| <i>Sargur Srihari, Anantharaman Ganesh, Catalin Tomai, Yong-Chul Shin, and Chen Huang</i> | |

| | |
|-------------------------------------------------------------|-----|
| Word-Wise Script Identification from Indian Documents | 310 |
| <i>Suranjit Sinha, Umapada Pal, and B.B. Chaudhuri</i> | |

| | |
|----------------------------------------------------|-----|
| Recognizing Freeform Digital Ink Annotations | 322 |
| <i>Michael Shilman and Zile Wei</i> | |

| | |
|---------------------------------------------------------------------------------------------------|-----|
| Post-processing of Handwritten Pitman's Shorthand Using Unigram and Heuristic Approaches | 332 |
| <i>Swe Myo Htwe, Colin Higgins, Graham Leedham, and Ma Yang</i> | |

| | |
|---------------------------------------------------------------------------|-----|
| Multiscale Handwriting Characterization for Writers' Classification | 337 |
| <i>Véronique Eglin, Stéphane Bres, and Carlos Rivero</i> | |

Graphics Recognition

| | |
|------------------------------------------------------------------|-----|
| A Hybrid Approach to Detect Graphical Symbols in Documents | 342 |
| <i>Salvatore Tabbone, Laurent Wendling, and Daniel Zuwala</i> | |

| | |
|----------------------------------------------------|-----|
| Performance Evaluation of Symbol Recognition | 354 |
| <i>Ernest Valveny and Philippe Dosch</i> | |

| | |
|--------------------------------------------------------------------------------------------------------------------|-----|
| The Search for Genericity in Graphics Recognition Applications: Design Issues of the Qgar Software System | 366 |
| <i>Jan Rendek, Gérald Masini, Philippe Dosch, and Karl Tombre</i> | |

| | |
|----------------------------------------------------------------------|-----|
| Attributed Graph Matching Based Engineering Drawings Retrieval | 378 |
| <i>Rujie Liu, Takayuki Baba, and Daiki Masumoto</i> | |

| | |
|--------------------------------------------------------------------------------------------------------------------------------|-----|
| A Platform to Extract Knowledge from Graphic Documents. Application to an Architectural Sketch Understanding Scenario | 389 |
| <i>Gemma Sánchez, Ernest Valveny, Josep Lladós, Joan Mas, and Narcís Lozano</i> | |

Internet Documents

| | |
|---------------------------------------------------------------------------------------------------------------|-----|
| A Graph-Based Framework for Web Document Mining | 401 |
| <i>Adam Schenker, Horst Bunke, Mark Last, and Abraham Kandel</i> | |
| XML Documents Within a Legal Domain: Standards and Tools for the Italian Legislative Environment | 413 |
| <i>Carlo Biagioli, Enrico Francesconi, Pierluigi Spinosa, and Mirco Taddei</i> | |
| Rule-Based Structural Analysis of Web Pages | 425 |
| <i>Fabio Vitali, Angelo Di Iorio, and Elisa Ventura Campori</i> | |
| Extracting Table Information from the Web | 438 |
| <i>Yeon-Seok Kim and Kyong-Ho Lee</i> | |
| A Neural Network Classifier for Junk E-Mail | 442 |
| <i>Ian Stuart, Sung-Hyuk Cha, and Charles Tappert</i> | |

Document Analysis Systems

| | |
|-------------------------------------------------------------------------------------------------------------------------|-----|
| Results of a Study on Invoice-Reading Systems in Germany | 451 |
| <i>Bertin Klein, Stevan Agne, and Andreas Dengel</i> | |
| A Document Analysis System Based on Text Line Matching of Multiple OCR Outputs | 463 |
| <i>Yasuaki Nakano, Toshihiro Hananoi, Hidetoshi Miyao, Minoru Maruyama, and Ken-ichi Maruyama</i> | |
| DocMining: A Document Analysis System Builder | 472 |
| <i>Sébastien Adam, Maurizio Rigamonti, Eric Clavier, Éric Trupin, Jean-Marc Ogier, Karl Tombre, and Joël Gardes</i> | |
| Automatic Fax Routing | 484 |
| <i>Paul Viola, James Rinker, and Martin Law</i> | |

Applications

| | |
|---------------------------------------------------------------------------------------------------------------------------|-----|
| Contextual <i>Swarm</i> -Based Multi-layered Lattices: A New Architecture for Contextual Pattern Recognition | 496 |
| <i>David G. Elliman and Sherin M. Youssef</i> | |
| Natural Language Processing of Patents and Technical Documentation . . . | 508 |
| <i>Gaetano Cascini, Alessandro Fantechi, and Emilio Spinicci</i> | |
| Document Image Retrieval in a Question Answering System for Document Images | 521 |
| <i>Koichi Kise, Shota Fukushima, and Keinosuke Matsumoto</i> | |
| A Robust Braille Recognition System | 533 |
| <i>Apostolos Antonacopoulos and David Bridson</i> | |

| | |
|-------------------------------------------------------------------------------------------------------|-----|
| Document Image Watermarking Based on Weight-Invariant Partition Using Support Vector Machine | 546 |
| <i>Shiyan Hu</i> | |
| Video Degradation Model and Its Application to Character Recognition in e-Learning Videos | 555 |
| <i>Jun Sun, Yutaka Katsuyama, and Satoshi Naoi</i> | |
| Unity Is Strength: Coupling Media for Thematic Segmentation | 559 |
| <i>Dalila Mekhaldi, Denis Lalanne, and Rolf Ingold</i> | |
| Author Index | 563 |

Document Analysis Systems VI

6th International Workshop, DAS 2004, Florence, Italy,

September 8-10, 2004, Proceedings

Marinai, S.; Dengel, A. (Eds.)

2004, XII, 568 p., Softcover

ISBN: 978-3-540-23060-1