TABLE OF CONTENTS

VOLUME 1

Welcome from the Honourary Chair
Welcome from the General Chairs
Message from the PC Chairs
Conference Committees
Reviewers

KEYNOTE I

A View on the Past and Future of Character and Document Recognition .........................................................1

Hiromichi Fujisawa

SESSION A1: HANDWRITING RECOGNITION (1)

Using Hidden Markov Models as a Tool for Handwritten Text Line Segmentation ...........................................6

Florence Luthy, Tamas Varga, Horst Bunke

Character Stroke Extraction Based on B-spline Curve Matching by Constrained Alternating Optimization .................................................................................................................................11

Xiabi Liu, Yunde Jia

Non-Uniform Slant Correction for Handwritten Text Line Recognition .................................................................16

Roman Bertolami, Seiichi Uchida, Matthias Zimmermann, Horst Bunke

A Method of Annotation Extraction from Paper Documents Using Alignment Based on Local Arrangements of Feature Points .................................................................................................21

Tomohiro Nakai, Koichi Kise, Masakazu Iwamura

Learning to Group Text Lines and Regions in Freeform Handwritten Notes .......................................................26

Ming Ye, Paul Viola, Sashi Raghupathy, Herry Sutanto, Chengyang Li

Character-Stroke Detection for Text-Localization and Extraction .............................................................................31

Krishna Subramanian, Prem Natarajan, Michael Decerbo, David Castañón

SESSION A2: ON-LINE HANDWRITING RECOGNITION

Kernel Modified Quadratic Discriminant Function for Online Handwritten Chinese Characters Recognition .................................................................36

Duanduan Yang, Lianwen Jin

A Minimax Classification Approach to HMM-Based Online Handwritten Chinese Character Recognition Robust Against Affine Distortions ..................................................................................41

Qiang Huo, Tingting He

Online Handwritten Japanese Character String Recognition Incorporating Geometric Context .................................................................................................................................................................................................46

Xiang-Dong Zhou, Jin-Lun Yu, Cheng-Lin Liu, Takeshi Nagasaki, Katsumi Marukawa
Minimum Error Discriminative Training for Radical-based Online Chinese Handwriting Recognition
Yaodong Zhang, Peng Liu, Frank K. Soong

Direction Code Based Features for Recognition of Online Handwritten Characters of Bangla
Ujjwal Bhattacharya, Bikash K. Gupta, Swapan K. Parui

HMM-based Online Handwriting Recognition System for Telugu Symbols
V. Jagadeesh Babu, L. Prasanth, R. Raghunath Sharma, G. V. Prabhakara Rao, A. Bharath

SESSION A3: DOCUMENT RETRIEVAL

A Fast Keyword-Spotting Technique
Linlin Li, Shijian Lu, Chew Lim Tan

Locality Sensitive Pseudo-Code for Document Images
Kengo Terasawa, Yuzuru Tanaka

XML Data Representation in Document Image Analysis
Abdel Belda, Ingrid Falk, Yves Rangoni

Language Models for Handwritten Short Message Services
Emmanuel Prochasson, Christian Viard-Gaudin, Emmanuel Morin

Vector Model Based Indexing and Retrieval of Handwritten Medical Forms
Huaigu Cao, Venu Govindaraju

Document Content Inventory & Retrieval
Michael A. Moll, Henry S. Baird

POSTER SESSION 1

Strategies for Large Handwritten Farsi/Arabic Lexicon Reduction
Saeed Mozaffari, Karim Faez, Volker Märgner, Haikal El Abed

Fusing Asynchronous Feature Streams for On-line Writer Identification
Andreas Schlapbach, Horst Bunke

Writer Identification in Handwritten Documents
Imran Ahmed Siddiqui, Nicole Vincent

Robust Document Warping with Interpolated Vector Fields
David C. Schneider, Marco Block, Raúl Rojas

Simultaneous Layout Style and Logical Entity Recognition in a Heterogeneous Collection of Documents
Siyuan Chen, Song Mao, George R. Thoma

Learning of Pattern-Based Rules for Document Classification
Andreas Dengel

A MSD-HMM Approach to Pen Trajectory Modeling for Online Handwriting Recognition
Lei Ma, Yi-Jian Wu, Peng Liu, Frank Soong

HMM-Based Recognizer with Segmentation-free Strategy for Unconstrained Chinese Handwritten Text
Tong-Hua Su, Tian-Wen Zhang, Hu-Jie Huang, Yu Zhou
Text Line Segmentation of Historical Arabic Documents ................................................................. 136
Abderrazak Zahour, Laurence Likforman-Sulem, Wafa Boussalaa, Bruno Taconet

Document Images Retrieval Based on Multiple Features Combination .............................................. 141
Gaofeng Meng, Nanning Zheng, Yonghong Song, Yuanlin Zhang

How to Reduce the Size of Bank Check Image Archive? ................................................................. 146
Vladimir Shapiro

A Human-Centric Off-Line Signature Verification System ................................................................. 151
Hanno Coetzer, Robert Sabourin

A Model-based Book Dewarping Method to Handle 2D Images Captured by a Digital Camera .......... 156
Minghui Wu, Rongfeng Li, Bin Fu, Wenzin Li, Zhuoqun Xu

Ink Annotations and their Anchoring in Heterogeneous Digital Documents ..................................... 161
Xin Wang, Sashi Raghupathy

A Method for the Synthesis of Dynamic Biometric Signature Data ................................................ 166
C. Rabasse, R.M. Guest, M.C. Fairhurst

Recognition of Broken Characters from Historical Printed Books Using Dynamic Bayesian Networks .......................................................... 171
Laurence Likforman-Sulem, Marc Sigelle

An Efficient Word Segmentation Technique for Historical and Degraded Machine-Printed Documents .......................................................... 176
M. Makridis, N. Nikolaou, B. Gatos

Circular Noises Removal from Scanned Document Images ............................................................... 181
Gaofeng Meng, Nanning Zheng, Yuanlin Zhang, Yonghong Song

Automatic Extraction of Data from 2-D Plots in Documents ........................................................... 186
Xiaonan Lu, James Z. Wang, Prasenjit Mitra, C. Lee Giles

An Efficient Thresholding Algorithm for Brazilian Bank Checks .................................................... 191
Carlos A.B.Mello, Byron L.D.Bezerra, Cleber Zanchettin, Valmir Macário

Recognition of Fragmented Characters Using Multiple Feature-Subset Classifiers .......................... 196
Chien-Hsing Chou, Chien-Yang Guo, Fu Chang

Minoru Yoshida, Hiroshi Nakagawa

Cryptogram Decoding for OCR using Numerization Strings ......................................................... 206
Gary Huang, Erik Learned-Miller, Andrew McCallum

Knowledge-based Recognition of Utility Map Sub-diagrams ....................................................... 211
Simon J. Hickinbotham, Anthony G. Cohn

Synthesis of Chinese Character Using Affine Transformation ..................................................... 216
LianWen Jin, XiaoNa Zu

Adaptive Window Based Uneven Lighting Document Segmentation ............................................. 221
Guoqing Gu, Wenwen Han

A Review of Shape Descriptors for Document Analysis .................................................................. 225
O. Ramos Terrades, S. Tabbone, E. Valveny

A High Performance European OCR System ................................................................................. 230
Kai Wang, Qingren Wang

Automatic Detection of Document Script and Orientation ............................................................. 235
Shijian Lu, Chew Lim Tan
<table>
<thead>
<tr>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dynamic Handwritten Keyword Spotting based on the NSHP-HMM</td>
<td>240</td>
</tr>
<tr>
<td>Christophe Choisy</td>
<td></td>
</tr>
<tr>
<td>Contribution of Multiresolution Description for Archive Document Structure Recognition</td>
<td>245</td>
</tr>
<tr>
<td>Aurelie Lemaître, Jean Camillerapp, Bertrand Couasnon</td>
<td></td>
</tr>
<tr>
<td>Iterated Document Content Classification</td>
<td>250</td>
</tr>
<tr>
<td>Chang An, Henry S. Baird, Pingping Xiu</td>
<td></td>
</tr>
<tr>
<td>Human Image Preference and Document Degradation Models</td>
<td>255</td>
</tr>
<tr>
<td>Chris Hale, Elisa H. Barney Smith</td>
<td></td>
</tr>
<tr>
<td>Text and Layout Information Extraction from Document Files of Various Formats Based on the Analysis of Page Description Language</td>
<td>260</td>
</tr>
<tr>
<td>Takashi Hirano, Yuichi Okano, Yasuhiro Okada, Fumio Yoda</td>
<td></td>
</tr>
<tr>
<td>Curved Paper Rectification for Digital Camera Document Images by Shape from Parallel Geodesics Using Continuous Dynamic Programming</td>
<td>265</td>
</tr>
<tr>
<td>Katsuhito Fujimoto, Jun Sanb, Hiroaki Takebea, Misako Suwaa, Satoshi Naoi</td>
<td></td>
</tr>
<tr>
<td>Annotated Databases for the Recognition of Screen-Rendered Text</td>
<td>270</td>
</tr>
<tr>
<td>Steffen Wachenfeld, Hans-Ulrich Klein, Xiaoai Jiang</td>
<td></td>
</tr>
<tr>
<td>An EM Based Algorithm for Skew Detection</td>
<td>275</td>
</tr>
<tr>
<td>Amir Egozi, Itshak Dinstein</td>
<td></td>
</tr>
<tr>
<td>Automatic Writer Identification Based on a Velocity Profile Approach to Processing of Difficult Handwriting Styles</td>
<td>280</td>
</tr>
<tr>
<td>J. Chapram, M. Fairhurst</td>
<td></td>
</tr>
<tr>
<td>A Novel Binarization System for Degraded Document Images</td>
<td>285</td>
</tr>
<tr>
<td>Yan Xi, Youbin Chen, Qingmin Liao, Leung Winghong, Fung Shunning, Deng Jiangwen</td>
<td></td>
</tr>
<tr>
<td>A Suffix Tree Based Handwritten Chinese Address Recognition System</td>
<td>290</td>
</tr>
<tr>
<td>Yan Jiang, Xiaoying Ding, Zheng Ren</td>
<td></td>
</tr>
<tr>
<td>Language-Based Feature Extraction Using Template-Matching in Farsi/Arabic Handwritten Numeral Recognition</td>
<td>295</td>
</tr>
<tr>
<td>Majid Ziaratban, Karim Faez, Farhad Faradji</td>
<td></td>
</tr>
<tr>
<td>Recognition of Character Strings in Low-quality Images Using Character and Inter-character Space Patterns</td>
<td>300</td>
</tr>
<tr>
<td>Hiroaki Ishida, Tomokazu Takahashi, Ichiro Ide, Hiroshi Murase</td>
<td></td>
</tr>
<tr>
<td>Locating Charts from Scanned Document Pages</td>
<td>305</td>
</tr>
<tr>
<td>Weihtau Huan, g, Chew Lim Tan</td>
<td></td>
</tr>
<tr>
<td>Binarization of Badly Illuminated Document Images through Shading Estimation and Compensation</td>
<td>310</td>
</tr>
<tr>
<td>Shijian Lu, Chew Lim Tan</td>
<td></td>
</tr>
<tr>
<td>A Multi-Stream Approach to Off-Line Handwritten Word Recognition</td>
<td>315</td>
</tr>
<tr>
<td>Yousri Kessentini, Thierry Paquet, AbdelMajid Benhamadou</td>
<td></td>
</tr>
<tr>
<td>Finger-Written Chinese Characters Recognition using Hierarchical LDA</td>
<td>320</td>
</tr>
<tr>
<td>Dunanduan Yang, Lianwen Jin, Li-Xin Zhen, Jiang-Cheng Huang</td>
<td></td>
</tr>
<tr>
<td>A Case-based Reasoning Approach for Invoice Structure Extraction</td>
<td>325</td>
</tr>
<tr>
<td>Hatem Hamza, Yolande Bela d, Abdel Bela d</td>
<td></td>
</tr>
</tbody>
</table>
An Approach to Improve Accuracy Rate of On-line Signature Verification Systems of Different Sizes

Rodrigo S. A. Araujo, George D. C. Cavalcanti, Edson C. B. C. Filho

KEYNOTE II

Energy-Based Models in Document Recognition and Computer Vision

Yann LeCun, Sumit Chopra, Marc'Aurelio Ranzato, Fu-Jie Huang

SESSION B1: HANDWRITING RECOGNITION (2)

Learning Handwritten Digit Recognition by the Max-Min Posterior Pseudo-Probabilities Method

Xuefeng Chen, Xiabi Liu, Yunde Jia

Transform-Invariance in Local Averaging Classifier for Handwritten Digit Pattern Recognition

Seiji Hotta

Identification of Latin-based Languages through Character Stroke Categorization

Shijian Lu, Linlin Li, Chew Lim Tan

Layout Analysis of Handwritten Historical Documents for Searching the Archive of the Cabinet of the Dutch Queen

Marius Bulacu, Ruiger van Koert, Lambert Schomaker, Tijn van der Zant

Indexing Historical Documents by Word Shape Signatures

Josep Llados, Gemma Sanchez

SESSION B2: ON-LINE HANDWRITING RECOGNITION (2)

A Novel Approach to On-Line Handwriting Recognition Based on Bidirectional Long Short-Term Memory Networks

Marcus Liwicki, Alex Graves, Horst Bunke, Jurgen Schmidhuber

Combining On-Line and Off-Line Systems for Handwriting Recognition

Marcus Liwicki, Horst Bunke

Text/Non-text Ink Stroke Classification in Japanese Handwriting Based on Markov Random Fields

Xiang-Dong Zhou, Cheng-Lin Liu

Use of a Confusion Network to Detect and Correct Errors in an On-Line Handwritten Sentence Recognition System

Solen Quiniou, Eric Anquetil

Redundant Bit Vectors for Robust Indexing and Retrieval of Electronic Ink

Kumar Chellapilla, John Platt

SESSION B3: PAGE SEGMENTATION

Robust Page Segmentation Based on Smearing and Error Correction Unifying Top-down and Bottom-up Approaches

Huaigu Cao, Rohit Prasad, Prem Natarajan, Ehry MacRostie
Decompose Document Image Using Integer Linear Programming ............................................................ 395
Dashan Gao, Yizhou Wang, Haitham Hindi, Minh Do

Character Line Segmentation Based on Feature Clustering ................................................................. 400
Yan Xi, Youbin Chen, Qingmin Liao, Leung Winghong, Fung Shunming, Deng Jiangwen

Document Image Segmentation Using a 2D Conditional Random Field Model ........................................... 405
Stéphane Nicolas, Julien Dardenne, Thierry Paquet, Laurent Heutte

Text Segmentation from Complex Background Using Sparse Representations ........................................... 410
W.M. Pan, T.D. Bui, C.Y. Suen

SESSION C1: CLASSIFIER

Exploiting Fisher Kernels in Decoding Severely Noisy Document Images ...................................................... 415
Jindong Chen, Yizhou Wang

K-Nearest Oracle for Dynamic Ensemble Selection ...................................................................................... 420
Albert Hung-Ren Ko, Robert Sabourin, Alceu de Souza Britto Jr.

Multi-Objective Optimization for SVM Model Selection ............................................................................. 425
C. Chatelain, S. Adam, Y. Lecourtier, L. Heutte, T. Paquet

A Classifier of Similar Characters using Compound Mahalanobis Function based on Difference Subspace ................................................................................................................... 430
Junichi Hirayama, Hidehisa Nakayama, Nei Kato

Extraction of Embedded Class Information from Universal Character Pattern ............................................. 435
Seiichi Uchida, Megumi Sakai, Masakazu Iwamura, Shinichiro Omachi, Koichi Kise

SESSION C2: ON-LINE HANDWRITING APPLICATIONS

Context-Driven Constraint Multiset Grammars with Incremental Parsing for On-line Structured Document Interpretation ........................................................................................................... 440
Sebastien Mace, Eric Anquetil

On-Line Handwritten Text Line Detection Using Dynamic Programming ..................................................... 445
Marcus Livicki, Emanuel Indermühle, Horst Bunke

An Incremental On-line Parsing Algorithm for Recognizing Sketching Diagrams ........................................ 450
Joan Mas, Gemma Sanchez, Josep Llados, Bart Lamirou

New Aspects of InkML for Pen-Based Computing ...................................................................................... 455
Stephen M. Watt

Frechet Distance based Approach for Searching Online Handwritten Documents ....................................... 459
R. Sriraghavendra, K. Karthik, Chiranjib Bhattacharyya

SESSION C3: COLOR ANALYSIS AND PERFORMANCE EVALUATION

Color Document Synthesis as a Compression Strategy .................................................................................. 464
João Marcelo Monte da Silva, Rafael Dueire Lins

Fast Selection of Small and Precise Candidate Sets from Dictionaries for Text Correction Tasks .................. 469
Stoyan Mihov, Peter Mitankin, Klaus U. Schulz
Automatic Ground-truth Generation for Document Image Analysis and Understanding ..................474
Pierre Héroux, Eugen Barbu, Sébastien Adam, Éric Trupin

New Metrics for Evaluating Performance in Document Analysis Tasks - Application to the Table Case - .................................................................479
Ana Costa e Silva

Identification of Non-Black Inks using HSV Colour Space ..................................................484
Haritha Dasari, Chakravarthy Bhagvati

SESSION D1: ASIAN CHARACTER RECOGNITION

Curvelet–based Multi SVM Recognizer for Offline Handwritten Bangla: A Major Indian Script .................................................................489
B.B. Chaudhuri, A. Majumdar

Off-Line Handwritten Character Recognition of Devanagari Script ..................................494
U. Pal, N. Sharma, T. Wakabayashi, F. Kimura

Modular Approach to Recognition of Strokes in Telugu Script .........................................499
Anitha Javeraman, C. Chandra Sekhar, V. Srinivasa Chakravarthy

Hidden Markov Models for Online Handwritten TamilWord Recognition .........................504
A. Bharath, Sriganesh Madhvanath

A Two Stage Recognition Scheme for Handwritten Tamil Characters ................................509

SESSION D2: GRAPHICS RECOGNITION

A New Syntactic Approach to Graphic Symbol Recognition .................................................514
Yu Yajie, Wan Zhang, Liu Wenyin

Extraction of Vectorized Graphical Information from Scientific Chart Images ....................519
Ruizhe Liu, Weihua Huang, Chew Lim Tan

Robust Circle Detection .......................................................................................................524
Bart Lamiroy, Olivier Gaucher, Laurent Fritz

Toponym Recognition in Scanned Color Topographic Maps .............................................529
Joachim Pouderoux, Jean-Christophe Gonzato, Aurelien Pereira, Pascal Guitton

A New Vectorial Signature for Quick Symbol Indexing, Filtering and Recognition .............534
Wan Zhang, Liu Wenyin

SESSION D3: IMAGE PREPROCESSING AND IMAGE CLASSIFICATION

PDE-based Enhancement of Low Quality Documents .........................................................539
Ifeoma Nwogu, Zhixin Shi, Venu Govindaraju

Non-Local Text Image Reconstruction ..............................................................................544
Heip Quang Luong, Wilfried Philips

SVM Based Scheme for Thai and English Script Identification ......................................549
S. Chanda, Oriol Ramos Terrades, U. Pal

A Hybrid System for Robust Recognition of Ethiopic Script ........................................554
Yaregal Assabie, Josef Bigun
POSTER SESSION II

Assessing and Improving the Quality of Document Images Acquired with Portable Digital Cameras ................................................................. 564
Rafael Dueire Lins, Gabriel Pereira e Silva, André Richardson Gomes e Silva

A Multi-stage Strategy to Perspective Rectification for Mobile Phone Camera-based Document Images ............................................................... 569
Xu-Cheng Yin, Jun Sun, Satoshi Naoi, Katsuhiro Fujimoto, Hiroaki Takebe, Yusaku Fujii, Koji Kurokawa

Degraded Character Recognition by Complementary Classifiers Combination ................................................................. 574
Jun Sun, Kaizhu Huang, Yoshinobu Hotta, Katsuhiro Fujimoto, Satoshi Naoi

Middle Zone Component Extraction and Recognition of Telugu Document Image ................................................................. 579
Pratap Reddy L, Satyaprasad L

An SVM-Based High-accurate Recognition Approach for Handwritten Numerals by Using Difference Features ................................................................. 584
Kaizhu Huang, Jun Sun, Yoshinobu Hotta, Katsuhiro Fujimoto, Satoshi Naoi

Retrieval of Handwritten Lines in Historical Documents ................................................................. 589
L.R.B. Schomaker

Text Line Detection in Unconstrained Handwritten Documents Using a Block-Based Hough Transform Approach ................................................................. 594
G. Louloudis, B. Gatos, C. Halatsis

Systematic Multi-Path HMM Topology Design for Online Handwriting Recognition of East Asian Characters ................................................................. 599
Shi Han, Ming Chang, Yu Zou, Xinjian Chen, Dongmei Zhang

Bibliographic Meta-Data Extraction Using Probabilistic Finite State Transducers ................................................................. 604
Martin Kramer, Hagen Kaprykowsky, Daniel Keysers, Thomas Breuel

Perceptive Vision for Headline Localisation in Bangla Handwritten Text Recognition ................................................................. 609
Aurelie Lemaitre, Bidyut Baran Chaudhuri, Bertrand Couasnon

Segmentation of Text and Graphics from Document Images ................................................................. 614
S.P. Chowdhury, S. Mandal, A.K. Das, Bhabatosh Chanda

Layout Based Information Extraction from HTML Documents ................................................................. 619
Radek Burget

An Overview of the Tesseract OCR Engine ................................................................. 624
Ray Smith

Symbol Recognition using a 2-class hierarchical model of Choquet Integrals ................................................................. 629
Jan Rendek, Laurent Wendling

Deriving Symbol Dependent Edit Weights for Text Correction - the Use of Error Dictionaries ................................................................. 634
Christoph Ringlstep, Ulrich Reffle, Annete Gotscharek, Klaus U. Schulz

Arabic Handwriting Recognition Using Variable Duration HMM ................................................................. 639
Amlan Kundu, Tom Hines, Jon Phillips, Benjamin D. Huyck, Linda C. Van Guilder
## Curvelets Based Queries for CBIR Application in Handwriting Collections

_G. Joutel, V. Eglin, S. Bres, H. Emptoz_ 644

## Two Approaches for Arabic Script Recognition-based Segmentation Using the Hough Transform

_Sofiène Touj, Najoua Essoukri Ben Amara, Hamid Amiri_ 649

## A General Method of Segmentation-Recognition Collaboration Applied to Pairs of Touching and Overlapping Symbols

_Christophe Renaudin, Yann Ricquebourg, Jean Camilleri_ 654

## Extracting “Carbon Copy” Names and Organizations From a Heterogeneous Document Collection

_Kazem Taghva, Russell Beckley, Jeffrey Coombs_ 659

## An Analytic Word Recognition Algorithm Using a Posteriori Probability

_Tomoyuki Hamamura, Takuma Akagi, Bunpei Irie_ 664

## Predictive DP Matching for On-Line Character Recognition

_Daiki Baba, Seiichi Uchida, Hiroaki Sakoe_ 669

## Matching Local Descriptors for Image Identification on Cultural Databases

_Eduardo Valle, Matthieu Cord, Sylvie Philipp-Foliguet_ 674

## Fast and Accurate Detection of Document Skew and Orientation

_Shijian Lu, Jie Wang, Chew Lim Tan_ 679

## Information Management System Using Structure Analysis of Paper/Electronic Documents and Its Applications

_Minenobu Seki, Masakazu Fujio, Takeshi Nagasaki, Hiroshi Shinjo, Katsumi Marukawa_ 684

## Mathpad: A Fuzzy Logic-Based Recognition System for Handwritten Mathematics

_John A. Fitzgerald, Franz Geiselbrechtinger, Tahar Kechadi_ 689

## Patra: A Novel Document Architecture for Integrating Handwriting with Audio-visual Information

_Gaurav Harit, Vinod Mankar, Santanu Chaudhury_ 694

## Integrated Segmentation and Recognition of Mixed Chinese/English Document

_Yong Xia, Bai-Hua Xiao, Chun-Heng Wang, Ru-Wei Dai_ 699

## Appearance Based Models in Document Script Identification

_T. N. Vikram, D. S. Guru_ 704

## Logical Structure Analysis for Form Images with Arbitrary Layout by Belief Propagation

_Akihiro Minagawa, Yusaku Fujii, Hiroaki Takebe, Katsuhito Fujimoto_ 709

## An Automatic Mathematical Expression Understanding System

_Yu-sheng Guo, Lei Huang, Chang-ping Liu, Xin Jiang_ 714

## Modified Dynamic Time Warping for Stroke-based On-line Signature Verification

_Won-Du Chang, Jungpil Shin_ 719

## On the Use of Context-Dependent Modeling Units for HMM-Based Offline Handwriting Recognition

_Gernot A. Fink, Thomas Plotz_ 724

## Off-line Signature Verification Using Enhanced Modified Direction Features in Conjunction with Neural Classifiers and Support Vector Machines

_Vu Nguyen, Michael Blumenstein, Vallipuram Mathukkumarasamy, Graham Leedham_ 729
A Multi-Agent System for Hand-drawn Diagram Recognition .......................................................... 734
Giovanni Casella, Vincenzo Deufemia, Viviana Mascardi

Enhancement of Old Manuscript Images.......................................................................................... 739
Avekash Gupta, Sunil Kumar, Rajat Gupta, Santanu Chaudhury, Shiv Dutt Joshi

Handwritten Numeral Recognition of Six Popular Indian Scripts ...................................................... 744
U. Pal, T. Wakabayashi, N. Sharma, F. Kimura

Text Input system using Online Overlapped Handwriting Recognition for Mobile Devices .............. 749
Yojiro Tonouchi, Akinori Kawamura

Machine Dating of Handwritten Manuscripts .................................................................................. 754
Utpal Garain, S. K Parui, T. Paquet, L. Heutte

Off-line Handwritten Signature GPDS-960 Corpus......................................................................... 759
J. Francisco Vargas, Miguel A. Ferrer, Carlos M. Travieso, Jesús B. Alonso

Text-Independent Writer Identification and Verification on Offline Arabic Handwriting ................. 764
Marius Bulacu, Lambert Schomaker, Axel Brink

A New Physically Motivated Warping Model for Form Drop-Out .................................................. 769
Guy Rosman, Asaf Tzadok, Doron Tal

Global Contour Lines Reconstruction in Topographic Maps .............................................................. 774
Joachim Pouderoux, Salvatore Spinello

Combination of OCR Engines for Page Segmentation Based on Performance Evaluation ............... 779
Miquel Ferrer, Ernest Valveny

An Indexing Method for Graphical Documents ............................................................................... 784
S. Tabbone, D. Zuwala

Document Processing Via Trained Morphological Operators ............................................................ 789
Nina S. T. Hirata

Content-level Annotation of Large Collection of Printed Document Images ...................................... 794
C. V. Jawahar, Anand Kumar

Bibliographic Attributes Extraction with Layer-upon-Layer Tagging ............................................. 799
Wei Wei, Irwin King, Jimmy Ho-Man Lee

Guido: a Musical Score Recognition System .................................................................................. 804
Mariusz Szwoch

A Coupled Mean Shift-Anisotropic Diffusion Approach for Document Image Segmentation and Restoration .......................................................... 809
Fadoua Drira, Frank LeBourgeois, Hubert Emptoz

KEYNOTE III

Google Book Search: Document Understanding on a Massive Scale .................................................. 814
Luc Vincent

SESSION E1: WRITER IDENTIFICATION AND VERIFICATION

Towards Explainable Writer Verification and Identification Using Vantage Writers .......................... 819
Axel Brink, Lambert Schomaker, Marius Bulacu
A New Method for Writer Identification and Verification Based on Farsi/Arabic Handwritten Texts
F. Shahabi nejad, M. Rahmati

Off-line Signature Verification System Performance against Image Acquisition Resolution
J. Francisco Vargas, Miguel A. Ferrer, Carlos M. Travieso, Jesús B. Alonso

Writer Identification using Steered Hermite Features and SVM
Asim Imdad, Stephane Bres, Veronigue Eglin, Hubert Empozt, Carlos Rivero-Moreno

On Computing Strength of Evidence for Writer Verification
Harish Srinivasan, Shrivardhan Kabra, Chen Huang, Sargur Srihari

SESSION E2: SYMBOL RECOGNITION

Mathematical Formula Recognition using 2D Grammars
Daniel Prusa, Vaclav Hlavac

A Unified Framework for Symbol Segmentation and Recognition of Handwritten Mathematical Expressions
Yu Shi, Hai Yang Li, Frank K. Soong

Hybrid Mathematical Symbol Recognition using Support Vector Machines
Birendra Keshari, Stephen M. Watt

Automatic Document Logo Detection
Guangyu Zhu, David Doermann

A Bayesian Network Approach to Mode Detection for Interactive Maps
Don J.M. Willems, Louis. G. Vuurpijl

SESSION E3: CAMERA AND VIDEO APPLICATIONS

Robust Binarization for Video Text Recognition
Zohra Saidane, Christophe Garcia

A Parametric Spectral-based Method for Verification of Text in Videos
Vassilis Papavassiliou, Themis Stafylakis, Vassilis Katsouros, George Carayannis

Camera-Based Graphical Symbol Detection
Marcal Rusinol, Josep Llados, Philippe Dosh

A Weighted Finite-State Framework for Correcting Errors in Natural Scene OCR
Richard Beaufort, Celine Mancas-Thillou

Exploiting Videotext “Events” for Improved Videotext Detection
Hrishikesh B. Aradhye, Gregory K. Myers

SESSION F1: CHINESE HANDWRITING RECOGNITION

Skew Detection for Chinese Handwriting by Horizontal Stroke Histogram
Tong-Hua Su, Tian-Wen Zhang, Hu-Jie Huang, Yu Zhou

LDA-Based Compound Distance for Handwritten Chinese Character Recognition
Tian-Fu Gao, Cheng-Lin Liu
Building Compact MQDF Classifier for Off-line Handwritten Chinese Characters by
Subspace Distribution Sharing ................................................................. 904
Teng Long, Lian-Wen Jin

Handwritten Chinese Character Recognition Using Modified LDA and Kernel FDA ................. 909
Duanduan Yang, Lianwen Jin

SESSION F2: LAYOUT ANALYSIS

Example-Based Logical Labeling of Document Title Page Images ........................................ 914
Joost van Beusekom, Daniel Keysers, Faisal Shafait, Thomas M. Breuel

A Data Mining Approach to Reading Order Detection .......................................................... 919
Michelangelo Ceci, Margherita Berardi, Giuseppe A. Porcelli, Donato Malerba

A General Approach for Partitioning Web Page Content Based on Geometric and Style
Information ........................................................................................................ 924
Hui Guo, Jalal Mahmud, Yevgen Borodin, Amanda Stent, I.V. Ramakrishnan

Searching for Tables in Digital Documents ...................................................................... 929
Ying Liu, Kun Bai, Prasenjit Mitra, C. Lee Giles

SESSION F3: APPLICATIONS

A Pen-based Interface for Real-time Document Edition ...................................................... 934
Jose A. Rodriguez, Gemma Sanchez, Josep Llados

Computer Assisted Transcription of Handwritten Text Images ........................................ 939
Alejandro H. Toselli, Veronica Romero, Enrique Vidal, Luis Rodriguez

A Method for Image Local-difference Visualization .......................................................... 944
Etienne Baudrier, Alain Riffaud

iGesture: A General Gesture Recognition Framework ....................................................... 949
Beat Signer, Ueli Kurmann, Moira C. Norrie

SESSION G1: ARABIC RECOGNITION

Combination of HMM-Based Classifiers for the Recognition of Arabic Handwritten
Words ...................................................................................................................... 954
Ramy Al-Hajj, Chafic Mokbel, Laurence Likforman-Sulem

A Hybrid Approach for Off-line Arabic Handwriting Recognition Based on a Planar
Hidden Markov Modeling ....................................................................................... 959
Sameh Masmoudi Touj, Najoua Essoukri Ben Amara, Hamid Amiri

Shape-based Alphabet for Off-line Arabic Handwriting Recognition ................................ 964
F. Menasri, N. Vincent, E. Augustin, M. Cheriet

Comparison of Different Preprocessing and Feature Extraction Methods for Offline
Recognition of Handwritten Arabic Words ............................................................. 969
Haikal El Abed, Volker Margner
SESSION G2: CAMERA-BASED SYSTEMS

Fast Lexicon-Based Scene Text Recognition with Sparse Belief Propagation ................................................. 974
Jerod J. Weinman, Erik Learned-Miller, Allen Hanson

Removing Shading Distortions in Camera-based Document Images Using Inpainting and Surface Fitting With Radial Basis Functions ................................................................. 979
Li Zhang, Andy M. Yip, Chew Lim Tan

Segmentation Based Recovery of Arbitrarily Warped Document Images ........................................................... 984
B. Gatos, I. Pratikakis, K. Nitrogiannis

Keyword Spotting and Retrieval of Document Images Captured by a Digital Camera ........................................ 989
Shijian Lu, Chew Lim Tan

SESSION G3: APPLICATIONS (2)

Spoken Handwriting Verification using Statistical Models .................................................................................... 994
Andreas Humm, Rolf Ingold, Jean Hennebert

An Embedded OCR Software Architecture for Enhancing Portability ................................................................. 999
Seho Kim, Jaehwa Park, Young-Bin Kwon

Generating Copybooks from Consistent Handwriting Styles ............................................................................. 1004
Ralph Nies, Louis Vuurpijl

POSTER SESSION III

A Feature based on Encoding the Relative Position of a Point in the Character for Online Handwritten Character Recognition ........................................................................................................ 1009
Dinesh M, Murali Krishna Sridhar

Implementation of Content Analysis System for Recognition of Journals’ Table of Contents ............................................................. 1013
Young-Bin Kwon, Jaehwa Park

Preliminary Experiments in Layout Analysis of Handwritten Letters Based on Textural and Spatial Information and a 2D Markovian Approach ........................................................... 1018
Melanie Lema tre, Emmanuelle Grosicki, Edouard Geoffrois, Francoise Preteux

Elastic Matching of Online Handwritten Tamil and Telugu Scripts Using Local Features .................................... 1023
L.Prasanth,V.Jagadeesh Babu, R. Raghunath Sharma, G.V.Prabhakara Rao, M. Dinesh

Usage-oriented Performance Evaluation for Text Localization Algorithms ......................................................... 1028
Yichao Ma, Chunheng Wang, Baihua Xiao, Ruwei Dai

An Effective and Practical Classifier Fusion Strategy for Improving Handwritten Character Recognition .................. 1033
Qiang Fu, X.Q.Ding, T.Zh.Li, C.S.Liu

Using Random Forests for Handwritten Digit Recognition .................................................................................. 1038
Simon Bernard, Laurent Heutte, Sebastien Adam

An approach for Multifont Arabic characters features Extraction based on Contourlet Transform .......................... 1043
Nadia Ben Amor, Najoua Essoukri Ben Amara

A Proposition of Retrieval tools for Historical Document Images libraries ............................................................ 1048
Nicholas Journet, Remy Mullot, Veronique Eglin, Jean-Yves Ramel
PRAAD: Preprocessing and Analysis Tool for Arabic Ancient Documents ...........................................1053
Wafa Boussellaa, Abderrazak Zahour, Bruno Taconet, Adel Alimi, Abdellatif Benabdelahif

The hOCR Microformat for OCR Workflow and Results .................................................................1058
Thomas M. Breuel

OCR Accuracy Improvement through a PDE-Based Approach .......................................................1063
Fadoua Drira, Frank LeBourgeois, Hubert Emptoz

An Efficient Feature Extraction and Dimensionality Reduction Scheme for Isolated Greek
Handwritten Character Recognition ....................................................................................................1068
G.Vamvakas, B.Gatos, S. Petridis, N.Stamatopoulos

Using Grammar-based Recognizers for Symbol Completion in Diagrammatic Sketches .................1073
Gennaro Costagliola, Vincenzo Deufemia, Michele Risi

A Visualization Tool to Improve the Performance of a Classifier Based on Hidden Markov
Models ................................................................................................................................................1078
Gleidson Pegoretti da Silva, Masaki Nakagawa

On the Use of Lexeme Features for Writer Verification ...............................................................1083
Anurag Bhardwaj, Abhishek Singh, Harish Srinivasan, Sargur Srihari

Incremental Learning of First Order Logic Theories for the Automatic Annotations of
Web Documents ..................................................................................................................................1088
Floriana Esposito, Stefano Ferilli, Nicola Di Mauro, Teresa M.A. Basile

Handwritten Word Recognition using Conditional Random Fields .................................................1093
Shravya Shetty, Harish Srinivasan, Sargur Srihari

Three Decision Levels Strategy for Arabic and Latin Texts Differentiation in Printed and
Handwritten Natures ..........................................................................................................................1098
Mohamed Ben Jlaiel, Slim Kanoun, Adel Mohamed Alimi, Rémy Mullot

Neural Network for the Recognition of Handwritten Tunisian City Names ..................................1103
Imen Ben Cheikh, Aref Kacem

On the effects of sampling rate and interpolation in HMM-based dynamic signature
verification ...........................................................................................................................................1108
M. Martinez-Diaz, J. Fierrez, M. R. Freire, J. Ortega-Garcia

Skeletonization by Gradient Regularization and Diffusion ..............................................................1113
F. Le Bourgeois, H. Emptoz

Use of Legal Amount to Confirm or Correct the Courtesy Amount on Farsi Bank Checks ..........1118
Majid Ziaratban, Karim Faez, Mehdi Ezoji

A colour document interpretation: Application to ancient cadastral maps ....................................1123
Romain Raveaux, Jean-Christophe Burie, Jean-Marc Ogier

Arabic Handwriting Texture Analysis for Writer Identification Using the DWT-Lifting
Scheme ...............................................................................................................................................1128
Sami Gazzah, Najoua Essoukri Ben Amara

Recognition, Understanding and Aestheticization of Freehand Drawing Flowcharts .................1133
Wioleta Szwoch

Table Recognition and Understanding from PDF Files .......................................................................1138
Tamir Hassan, Robert Baumgartner

Ink Discrimination based on Co-occurrence Analysis of Visible and Infrared Images ..................1143
Vasiliki Kokla, Alexandra Psarrou, Vassilis Konstantinou

Segmentation of Very Low Resolution Screen-Rendered Text .........................................................1148
Steffen Wachenfeld, Hans-Ulrich Klein, Stefan Fleischer, Xiaoyi Jiang
<table>
<thead>
<tr>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Quantile Linear Algorithm for Robust Binarization of Digitalized Letters</td>
<td>1153</td>
</tr>
<tr>
<td>Marte Ramirez, Ernesto Tapia, Marco Block, Raul Rojas</td>
<td></td>
</tr>
<tr>
<td>A Shared Parts Model for Document Image Recognition</td>
<td>1158</td>
</tr>
<tr>
<td>Mithun Das Gupta, Prateek Sarkar</td>
<td></td>
</tr>
<tr>
<td>Context-Sensitive Error Correction: Using Topic Models to Improve OCR</td>
<td>1163</td>
</tr>
<tr>
<td>Michael L. Wick, Michael G. Ross, Erik G. Learned-Miller</td>
<td></td>
</tr>
<tr>
<td>On the Applicability of Off-line Signatures to the Fuzzy Vault Construction</td>
<td>1168</td>
</tr>
<tr>
<td>Manuel R. Freir, Julian Fierrez, Marcos Martinez-Diaz, Javier Ortega-Garcia</td>
<td></td>
</tr>
<tr>
<td>Autonomous Text Capturing Robot Using Improved DCT Feature and Text Tracking</td>
<td>1173</td>
</tr>
<tr>
<td>Makoto Tanaka, Hideaki Goto</td>
<td></td>
</tr>
<tr>
<td>Aspects of Mathematical Expression Analysis in Arabic Handwriting</td>
<td>1178</td>
</tr>
<tr>
<td>Elena Smirnova, Stephen M. Watt</td>
<td></td>
</tr>
<tr>
<td>Online Handwritten Kanji Recognition Based on Inter-stroke Grammar</td>
<td>1183</td>
</tr>
<tr>
<td>Ikumi Ota, Ryo Yamamoto, Shinji Sako, Shigeki Sagayama</td>
<td></td>
</tr>
<tr>
<td>Physical Structure Segmentation with Projection Profile for Mathematic Formulae and Graphics in Academic Paper Images</td>
<td>1188</td>
</tr>
<tr>
<td>Tzu-Yuan Chang, Yusuke Takiguchi, Minoru Okada</td>
<td></td>
</tr>
<tr>
<td>Representing and Characterizing Handwritten Mathematical Symbols through Succinct Functional Approximation</td>
<td>1193</td>
</tr>
<tr>
<td>Bruce W. Char, Stephen M. Watt</td>
<td></td>
</tr>
<tr>
<td>Extracting Features of Paper-made Objects Recognized from Origami Books Based on Design Knowledge</td>
<td>1198</td>
</tr>
<tr>
<td>Hiroshi Shimamuki, Jien Kato, Toyohide Watanabe</td>
<td></td>
</tr>
<tr>
<td>Word Image Based Latent Semantic Indexing for Conceptual Querying in Document Image Databases</td>
<td>1203</td>
</tr>
<tr>
<td>Sameek Banerjee, Gaurav Harit, Santanu Chaudhury</td>
<td></td>
</tr>
<tr>
<td>A Sparse and Locally Shift Invariant Feature Extractor Applied to Document Images</td>
<td>1208</td>
</tr>
<tr>
<td>Marc’Aurelio Ranzato, Yann LeCun</td>
<td></td>
</tr>
<tr>
<td>A Novel Hierarchical Classification Scheme for Online Tamil Character Recognition</td>
<td>1213</td>
</tr>
<tr>
<td>Suresh Sundaram, A. G. Ramakrishnan</td>
<td></td>
</tr>
<tr>
<td>Extraction of Arabic Words from Complex Color Image</td>
<td>1218</td>
</tr>
<tr>
<td>Radwa Fathalla, Yasser El Sonbaty, Mohamed A. Ismail</td>
<td></td>
</tr>
<tr>
<td>Image Pixel Force Fields and Their Application for Color Map Vectorisation</td>
<td>1223</td>
</tr>
<tr>
<td>V. Bucha, S. Uchida, S. Ablameyko</td>
<td></td>
</tr>
<tr>
<td>New Strategy for the on Line Handwriting Modeling</td>
<td>1228</td>
</tr>
<tr>
<td>Houcine Boubaker, Monji Kherallah, Adel M. Alimi</td>
<td></td>
</tr>
<tr>
<td>On Using Classical Poetry Structure for Indian Language Post-Processing</td>
<td>1233</td>
</tr>
<tr>
<td>Anoop M. Namboodiri, P. J. Narayanan, C. V. Jawahar</td>
<td></td>
</tr>
<tr>
<td>On Segmentation of Documents in Complex Scripts</td>
<td>1238</td>
</tr>
<tr>
<td>K. S. Sesh Kumar, Sukesh Kumar, C. V. Jawahar</td>
<td></td>
</tr>
<tr>
<td>A Blind Indic Script Recognizer for Multi-script Documents</td>
<td>1243</td>
</tr>
<tr>
<td>Peeta Basa Patti, A G Ramakrishnan</td>
<td></td>
</tr>
<tr>
<td>Streaming-Archival InkML Conversion</td>
<td>1248</td>
</tr>
<tr>
<td>Birendra Keshari, Stephen M. Watt</td>
<td></td>
</tr>
</tbody>
</table>
Performance Analysis Framework for Layout Analysis Methods ................................................................. 1253
  A. Antonacopoulos, D. Bridson

Unsupervised Newspaper Segmentation Using Language Context ............................................................. 1258
  Ralph Furmaniak

KEYNOTE IV

Advances in Writer Identification and Verification ....................................................................................... 1263
  Lambert Schomaker

COMPETITION SESSION

ICDAR 2007 - Arabic Handwriting Recognition Competition ................................................................. 1269
  Volker Margner, Haikal El Abed

ICDAR2007 Page Segmentation Competition .............................................................................................. 1274
  A. Antonacopoulos, B. Gatos, D. Bridson

ICDAR2007 Handwriting Segmentation Contest .......................................................................................... 1279
  B. Gatos, A. Antonacopoulos, N. Stamatopoulos

Author Index