

Sunday, 12 February

6:00 - 8:00 pm Registration / Reception



Monday, 13 February

8:00 Registration

8:45 Opening Remarks

Session 1: Horst Bunke

9:00 1.1 Keynote address: Digital libraries, developing countries, and document analysis systems
Ian Witten

9:40 1.2 Retrieval from document image collections
A. Balasubramanian, Million Meshesha, C. V. Jawahar

10:00 1.3 A semi-automatic adaptive OCR for digital libraries
Sachin Rawat, K. S. Sesh Kumar, Million Meshesha, Indineel Deb Sikdar, A. Balasubramanian, C. V. Jawahar

10:20 **Group A: Poster Spotlight**
Tin Kam Ho

10:50 **Break / Posters / Demos**

A.1 HVS inspired system for script identification in Indian multi-script documents
Peeta Basa Pati, A G Ramakrishnan

A.2 Offline handwritten Arabic character segmentation with probabilistic model
Pingping Xiu, Liangrui Peng, Xiaoqing Ding, Hua Wang

A.3 Automatic keyword extraction from historical document image
Kengo Terasawa, Takeshi Nagasaki, Toshio Kawashima

A.4 Digitizing a million books: Challenges for document analysis
C. V. Jawahar, Pramod Sankar K, Vamshi Ambati

A.5 Demo: Arabic word-spotting,
Jim Yaghi, Larry Spitz

Session 2: Image Processing

Apostolos Antonacopoulos

11:20 2.1 Contribution to the discrimination of the medieval manuscript texts: Application in the palaeography
Ikram Moalla, Frank LeBourgeois, Hubert Emptoz, Adel M. Alimi

11:40 2.2 Restoring degraded document images using a recursive unsupervised classification technique
Fadoua Drira, Frank Le Bourgeois, Hubert Emptoz

12:00 2.3 Networked document imaging with normalization and optimization
Hirobumi Nishida

Monday, 13 February

12:20 2.4 Gray-scale thinning algorithm using local min/max operations
Kyoung Min Kim, Gwan Hee Kang, Myung Hyun Song, Buhm, Lee, Joong Jo Park, Ching Y. Suen

12:40 Lunch

Session 3: Handwriting 1

Thomas Breuel

1:40 3.1 Automated scoring of handwritten essays based on latent semantic analysis
Sargur Srihari, Jim Collins, Rohini Srihari, Pavithra Babu, Harish Srinivasan

2:00 3.2 Aligning transcripts to automatically segmented handwritten manuscripts
J. Rothfeder, R. Manmatha, T. M. Rath

2:20 3.3 Virtual example synthesis based on PCA for off-line handwritten character recognition
Hidetoshi Miyao, Minoru Maruyama

2:40 3.4 Extraction of handwritten text from carbon copy medical form images
Robert Milewski, Venu Govindaraju

3:00 **Group B: Poster Spotlight**
Koichi Kise

3:30 **Break / Posters / Demos**

B.1 Reconstruction of orthogonal polygonal lines
Alexander Gribov, Eugene Bodansky

B.2 The restoration of camera documents through image segmentation
Lu Shijian, Chew Lim Tan

B.3 The impact of OCR accuracy and feature transformation on automatic text classification
Mayo Murata, Lazaro S.P. Busagala, Wataru Ohyama, Teisushi Wakabayashi, Fumitaka Kimura

B.4 Toward file consolidation by document categorization
A. Belaid, A Alusse

B.5 Demo: MyReader
Phillip Pearson, Paul Seakins, David Lee, Nigel Tucker, Darryl Sherwood

Session 4: Document Structure and Format

Daniel Lopresti

4:00 4.1 Document logical structure analysis based on perceptive cycles
Yves Rangoni, Abdel Belaid

4:20 4.2 A system for converting PDF documents into structured XML format
Hervé Déjean, Jean-Luc Meunier

4:40 4.3 CDF: A canonical and structured document format
Jean-Luc Bloechle, Maurizio Rigamonti, Denis Lalanne, Rolf Ingold

5:00 4.4 Structural analysis of mathematical formulae with verification based on formula description grammar
Seiichi Toyota, Seiichi Uchida, Masakazu Suzuki

5:20 Adjoin

Tuesday, 14 February

Session 5: Larry Spitz

8:30 5.1 Keynote address: Smart bombs and reading machines for the blind
Jim Fruchterman

9:10 5.2 Notes on contemporary table recognition
David W. Embley, Daniel Lopresti, George Nagy

9:30 5.3 Handwritten artefact identification method in table interpretation without using knowledge
Luiz Antônio Pereira Neves, João Marques de Carvalho, Jacques Facon, Flávio Bortolozzi, Sérgio Aparecido Ignácio

9:50 **Group C: Poster Spotlight**
Yasuaki Nakano

10:20 **Break / Posters / Demos**

C.1 A method for symbol spotting in graphical documents
Daniel Zuwala, Salvatore Tabbone

C.2 Groove extraction of phonographic records
Sylvain Stotzer, Ottar Johnsen, Frédéric Bapst, Rolf Ingold

C.3 Use of affine invariants in locally likely arrangement hashing for camera-based document image retrieval
Tomohiro Nakai, Koichi Kise, Masakazu Iwamura

C.4 Robust Chinese character recognition by selection of binary-based and grayscale-based classifier
Yoshinobu Hotta, Jun Sun, Yutaka Katsuyama, Satoshi Naoi

C.5 Segmentation-driven recognition applied to numerical field extraction from handwritten incoming mail documents
C. Chatelain, L. Heutte, T. Paquet

C.6 Demo: Tin Kam Ho

Session 6: Handwriting 2 - Venu Govindaraju

10:50 6.1 Writer identification for smart meeting room systems
Marcus Liwicki, Andreas Schlapbach, Horst Bunke, Samy Bengio, Johnny Mariéthoz, Jonas Richiardi

11:10 6.2 Extraction and analysis of document examiner
Vladimir Pervouchine, Graham Leedham

11:30 6.3 Segmentation of on-line handwritten Japanese text using SVM for improving text recognition
Bilan Zhu, Masaki Nakagawa

11:50 6.4 An OCR based Chinese handwritten address reading system
Yan Jiang, Xiaoqing Ding, Qiang Fu, Zheng Ren

12:10 Lunch

Session 7: Language and Script Identification

Andreas Dengel

1:10 7.1 Language identification in degraded and distorted document images
Lu Shijian, Chew Lim Tan

1:30 7.2 Bangla/English script identification based on analysis of connected component profiles
Lijun Zhou, Yue Lu, Chew Lim Tan

1:50 7.3 Script identification from Indian documents
Gopal Datt Joshi, Saurabh Garg, Jayanthi Sivaswamy



Tuesday, 14 February

- 2:10** 7.4 Finding the best-fit bounding-boxes
Bo Yuan, Leong Keong Kwoh, Chew Lim Tan

Session 8: Discussion Groups - Henry Baird

2:30 Discussion Group Pitch

3:00 Break

3:30 Discussion groups

5:00 Adjourn

6:30 Buses leave for Banquet



Wednesday, 15 February

Session 9: Systems and Performance Evaluation David Doermann

- 8:30** 9.1 Towards versatile document analysis systems
Henry S. Baird, Matthew R. Casey

- 8:50** 9.2 Exploratory analysis system for semi-structured engineering logs
Michael Flaster, Bruce Hillyer, Tin Kam Ho

- 9:10** 9.3 Ground truth for layout analysis performance evaluation
A. Antonacopoulos, D. Karatzas, D. Bridson

- 9:30** 9.4 Benchmarking of invoice analysis systems
Bertin Klein, Stefan Agne, Andreas Dengel

- 9:50** 9.5 Semi-automatic ground truth generation for chart image recognition
Li Yang, Weihua Huang, Chew Lim Tan

10:10 **Group D: Poster Spotlight**
Simone Marinai

10:40 Break / Posters / Demos

- D.1 Performance evaluation of text detection and tracking in video
Vasant Manohar, Padmanabhan Soundararajan, Matthew Boonstra, Harish Raju, Dmitry Goldgof, Rangachar Kasturi
- D.2 Automatic assembling of cadastral maps based on generalized Hough transformation
Fei Liu, Wataru Ohyama, Tetsushi Wakabayashi, Fumitaka Kimura
- D.3 A few steps towards on-the-fly symbol recognition with relevance feedback
Jan Rendek, Bart Lamiroy, Karl Tombre
- D.4 The fuzzy-spatial descriptor for the online graphic recognition: overlapping matrix algorithm
Noorazrin Zakaria, Jean Marc Ogier, Josep Llados
- D.5 Demo: Exbiblio
Martin King

Wednesday, 15 February

11:20 **Group E: Extended Abstracts**
R. Manmatha

- E.1 Objective quality measurement for geometric document image restoration
Christoph H. Lampert, Thomas M. Breuel
- E.2 The component-linking tree revisited
Bo Yuan, Leong Keong Kwoh, Chew Lim Tan
- E.3 A platform for web-based OCR systems with server search function
Hideaki Goto
- E.4 A table evaluation method for plain text documents
Vanessa Long, Steve Cassidy, Robert Dale
- E.5 Combined spam filter with bayesian filter and SVM filter
Ayahiko Niimi, Hirofumi Inomata, Masaki Miyamoto, Osamu Konishi
- E.6 Investigation of singular spectrum analysis and machine learning for road sign location
Kyoko Fukuda, Phillip Pearson
- E.7 Round-trip HTML rendering and analysis for testing, indexing, and security
Thomas Breuel, Daniel Keysers

12:00 Lunch

Session 10: Retrieval and Segmentation
Rangachar Kasturi

- 1:00** 10.1 Efficient word retrieval by means of SOM clustering and PCA
Simone Marinai, Stefano Faini, Emanuele Marino, Giovanni Soda

- 1:20** 10.2 The effects of OCR error on the extraction of private information
Kazem Taghva, Russell Beckley, Jeffrey Coombs

- 1:40** 10.3 Combining multiple classifiers for faster optical character recognition
Kumar Chellapilla, Michael Shilman, Patrice Simard

- 2:00** 10.4 Performance comparison of six algorithms for page segmentation
Faisal Shafait, Daniel Keysers, Thomas M. Breuel

2:20 Break

2:50 **Session 11: Reports of discussion groups**
Henry Baird

3:35 Concluding remarks

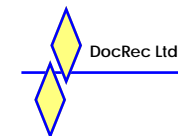


Seventh IAPR Workshop on Document Analysis Systems

13 - 15 February 2006

Workshop Co-Chairs:
Larry Spitz, DocRec Ltd
Horst Bunke, Universität Bern

Sponsored by:



Humanware™

HITACHI
Inspire the Next

SIEMENS

