

12:00–1:00 LUNCH (RM2101-2105)

1:00–2:00 **KEYNOTE SPEECH: EDUCE: Enhanced Digital Unwrapping for Conservation and Exploration of Inaccessible Texts**, W. BRENT SEALES, Univ. of Kentucky

2:15–3:35 **Session 7 – Document Analysis Applications, Chair: Henry Baird, Lehigh University**

GOVERNMENT BRIEFING: Supporting Document Analysis Research at DARPA, J. OliveDARPA

Automated Degraded Image Enhancement Program (DIEP™), G. Boutros CiyaSoft Corp.

A Document Triage System for Army Applications, F. Fisher, K. Marcus, R. Chang and J. Turner ARL

Multi-Language Handwriting Derived Biometric Identification, D. Gantz, J. Miller and M.A. Walch* George Mason Univ. and Gannon

Tradeoff Studies about Storage and Retrieval Efficiency of Boundary Data Representations for LLS, TIGER and DLG Data Structures, D. Clutter and P. Bajcsy NCSA, Univ. Illinois

3:35–4:35 **Demos/Abstracts** (RM 2104-2100)
ABBYY Software House, OCR, Forms Processing and Data Capture Technology Review, Chris Riley, ABBYY, USA

- **VERSUS A Middle Eastern Language OCR**, S.G. Schlosser, NovoDynamics, Inc.
- **IDG: A Business Information Extraction, Management, and Routing Front-End for Content Management Systems**, V. Krishna, S. Srinivasan, N. Boyette, I. Cheng, J. Kreulen and T. Kanungo, IBM Almaden
- *Fast Skew and Slant Correction for Arabic Written Word or Line*, E.M. Zaki and M. El-Adawi, Sakhr Software USA, Inc.
- *Pictographic Recognition Technique Applied to Distinctive Characteristics of Handwritten Arabic Text*, M. Walch and D. Gantz*, Gannon and George Mason Univ*
- **Embedded Translation Layer and Cross-Language Entity Extraction**, M. Shalev, *Encyclopedia Britannica*.
- **DOCLIB: A Software Library for Document Processing**, G. Zhu, K. Chen, S. Jaeger, UMCP
- **GEDI: A Ground-Truth Editor and Document Interface**, G. Zhu, K. Chen, S. Jaeger, UMCP

4:30–7:30 DEMOS/POSTERS RECEPTION
(FOUNDERS ROOM)

FRIDAY — NOVEMBER 4, 2005

8:30 AM BREAKFAST (2ND FLOOR LANDING)

9:30–12:00 **Workshop Evaluation of Document Image Processing Technologies** (RM 2104-2100)

Issues with Automatic OCR Evaluation and Metrics, K. Concepcion MITRE

Building Arabic Language Training and Testing Sets, M. Walch Gannon

Performance Evaluation of Multilingual Document Exploitation Systems, S. Schlosser NovoDynamics, Inc.

Ground Truth Representation Used in Testing and Optimization of the Optical Word Recognition System, M. Ladwig Northrop Grumman

Evaluation with Informational Confidence, S. Jaeger and D. Doermann UMCP

Metrics for Word Spotting, S. Srihari. CEDAR

12:00-1:00 PM WORKING LUNCH
(RM 2104-2100)

Performance Evaluation for Text Processing of Noisy Inputs, D. Lopresti Lehigh Univ.

VACE Video Text Recognition Evaluation Plans, J. Garofolo and R. Bowers NIST

Fast Skew and Slant Correction for Arabic Written Word on Line, M. Zaki and M. El-Adawi Sakhr, Helwan Univ.

Evaluation Issues in Image Refiner, K. Summers and E. Borovikov CACI

The Sporadic Nature of OCR Evaluations, T. Kanungo IBM Almaden

2:40-3:00 PM BREAK (2ND FLOOR LANDING)

3:00-4:00 *Panel Discussions*



FINAL PROGRAM

2005 Symposium on
Document Image
Understanding Technology

November 2-4, 2005

Inn and Conference Center
Marriott Hotel
Adelphi, Maryland

Post Symposium information will be available on the SDIUT Home Page
<http://lamp.cfar.umd.edu/sdiut05>
If you have any questions, please contact the Symposium Coordinators at:
Phone: (301) 405-6444 or 1767; Fax: (301) 314-2644
E-mail: sdiut05@umiacs.umd.edu
SDIUT05, UMIACS, University of Maryland, College Park, MD 20742

FINAL PROGRAM

WEDNESDAY — NOVEMBER 2, 2005

7:30 AM REGISTRATION/BREAKFAST
(2ND FLOOR LANDING)
9:00 AM WELCOME/INTRODUCTION
(RM 2104-2100)

9:15–10:35 **Session 1 - Document Analysis Systems – Chair: Daniel Lopresti, Lehigh University**

A Flexible Experimentation and Configuration Platform for Multilingual HMM OCR,
E. MacRostie and P. Natarajan BBN Technologies

Document Information Processing Towards Software and Test Collections, **G. Agam, S. Argamon, O. Frieder, D. Grossman and D. Lewis*** IIT and David D. Lewis Consulting*

DOCLIB: A Document Processing Research Tool, **K. Chen*, S. Jaeger, G. Zhu and D. Doermann** UMCP and BAH*

A Process Flow for Realizing High Accuracy for OCR Text, **K. Taghva, J. Borsack and T. Nartker** UNLV

10:35–11:00 BREAK (2ND FLOOR LANDING)

11:00–12:00 **Session 2 – Document Processing and Enhancement – Chair: Stefan Jaeger, UMCP**

GOVERNMENT BRIEFING: The NSA/HLT Program, B. Wheatley NSA

A Document Image Enhancement Module: Perspective Warp Correction, **C. Monnier and V. Ablavsky*, S. Holden and M. Snorrason** Charles River Analytics, Inc. and Boston Univ.*

Document Recognition and Translation with Handheld Mobile Devices, **E.D. Haritaoglu and I. Haritaoglu** Polar Rain, Inc.

12:00–1:00 LUNCH (CHESAPEAKE)

1:00–2:00 KEYNOTE SPEECH: Google Print – LUC VINCENT, Google, Inc.

2:15–3:35 **Session 3 – Processing of Arabic Documents – Chair: Venu Govindaraju, CEDAR, SUNY at Buffalo**

A System for Discriminating Handwriting from Machine Print on Noisy Arabic Documents
J.C. Femiani, M. Phielipp and A. Razdan Arizona State Univ.

NLP-Enhanced Exploitation of Degraded Arabic Texts, **M. Shalev** Encyclopedia Britannica

Initial Results in Offline Arabic Handwriting Recognition Using Large-Scale Geometric Features
I. Zavorin, E. Borovikov and M. Turner CACI

Classification of Machine Print and Handwriting in Mixed Arabic Documents, **K. Sridharan, F. Farooq and V. Govindaraju** CEDAR

3:35–3:55 BREAK (2ND FLOOR LANDING)

3:55–5:15 **Session 4 – Extraction and Information Retrieval – Chair: Kazem Taghva, Univ. of Nevada**

**GOVERNMENT BRIEFING: The Sequoyah Foreign Language Translation System,
E. Cerutti US Army**

Accurate Document Categorization of OCR Generated Text, **R. Price* and A. Zukas** SAIC and CAC*

Effect of Degraded Input on Statistical Machine Translation, **F. Farooq and Y. Al-Onaizan*** CEDAR, IBM TJ Watson

Mobile Interactive Support System, **G. Nagy and D. Lopresti*** RPI, Lehigh Univ. *

THURSDAY — NOVEMBER 3, 2005

8:00 AM REGISTRATION/BREAKFAST
(2ND FLOOR LANDING)

9:00–10:20 **Session 5 – Arabic Document Analysis – Chair: Premkumar Natarajan, BBN Technologies.**

GOVERNMENT BRIEFING: The LASER ACTD Program, L. Hernandez and M. Holland ARL

Handwritten Arabic Word Spotting Using the CEDARABIC Document Analysis System,
S. Srihari, H. Srinivasan, P. Babu and C. Bhole CEDAR

Word Spotting in Arabic Script Documents,
A.L. Spitz and Y. Yaghi DocRec, Ltd.

Challenges in Adapting Machine-Print Arabic OCR for Handwriting, **S.G. Schlosser and R.C. Vogt** NovoDynamics, Inc.

10:20–10:40 BREAK (2ND FLOOR LANDING)

10:40–12:00 **Session 6– Cross Domain Applications, Chair: Larry Spitz, DocRec Ltd.**

**GOVERNMENT BRIEFING: Government Needs at NMEC,
B. McGrath NMEC**

Robot Navigation Techniques for Engineering Drawing Analysis **T.C. Henderson and C. Xu** Univ. Utah

Perceptual Organization in Semantic Role Labeling, **P. Sarkar and E. Saund** PARC

Pictographic Recognition Technique Applied to Distinctive Characteristics of Handwritten Arabic Text, **M. Walch and D. Gantz*** Gannon and George Mason Univ.*