

# **Ninth IAPR International Workshop On Document Analysis Systems**

# **DAS 2010**

June 9 - 11, 2010 - Boston, MA

# **Workshop Program**



Supported by:

Raytheon



#### Welcome to DAS 2010

It is the pleasure of the organizing committee to welcome all participants to the 2010 IAPR workshop on Document Analysis and Recognition (DAS). This year's workshop is being held June 9-11<sup>th</sup> in Boston, Massachusetts, a location situated in the heart of beautiful New England in the northeastern United States. Boston has a rich history that dates back to the 1600's and was a major focal point in the history of the American Revolution and America's fight for independence. Over the years, Boston has developed into a center for industrial, academic and cultural excellence and draws millions of visitors each year.

DAS 2010 is the ninth workshop in a series. The first DAS was held in Kaiserslautern, Germany in 1994, and was followed by Malvern, USA (1996); Nagano, Japan (1998); Rio de Janeiro, Brazil (2000); Princeton, USA (2002); Florence, Italy (2004); Nelson, New Zealand (2006) and Nara Japan (2008). The DAS tradition is to bring together industry, academic and government researchers interested in many aspects of document analysis systems and to provide opportunities for fruitful interaction and collaboration. This year's workshop is organized as a three day, single track event, with oral and poster presentations, as well as working group discussions on the second and third afternoons. Special sessions on contributed datasets and a keynote talk on this same topic provide a compelling theme, and we hope this focus on this topic will help push the field toward greater sharing of data and accepted standards for evaluation.

This year, we received 91 submissions from 25 countries on 6 continents. The papers were reviewed by 45 members of our research community and an international program committee representing 16 different countries. The overall quality was excellent and we have chosen 28 full papers for oral presentation, 37 as poster papers, as well as 15 short papers that will be presented either as posters or in a special short oral session. In addition, six groups are scheduled to present live demos during the poster sessions.

Full papers underwent the standard peer review process and will appear in the official workshop proceedings to be published in the ACM International Conference Proceedings Series, available online as part of the ACM Digital Library. The short papers are included in the unofficial hardcopy proceedings distributed at the event as well as on the DAS 2010 website.

Professor George Nagy, *Professor of Computer Engineering at RPI in Troy, NY* and a founding father in the field of Document Image Analysis, is scheduled to kick-off the workshop with a keynote talk entitled "Document Systems Analysis: Testing, Testing, Testing". In the DAS tradition, we will also be presenting the IAPR Best Student Paper Award, as well as the IAPR Nakano Award for the best paper at this year's event. The latter was established in memory of the late Professor Yasuaki Nakano, the honorary chair of the DAS 2008 workshop and the general chair of DAS 1998.

Socially, the local organizers have arranged an exciting set of events including an opening reception, a duck boat tour of Boston and its harbor and a banquet at the "Top of the Hub" overlooking the city. We would like to thank Rohit Prasad, David Frampton, and Laura Stephens for their tireless work on managing the local arrangements, Srirangaraj Setlur for his work on the workshop web page and publications, and Jeanne Steinberg for here help in assembling the hardcopy proceedings. Their behind the scenes efforts are a key to the success of DAS2010. We also wish to take this opportunity to thank Raytheon BBN Technologies and Hitachi for their generous financial support.

We hope that this workshop reflects the excellence that has made DAS an important biennial event in our community. Please enjoy your time in Boston and your participation in DAS2010!

David Doermann, University of Maryland, College Park Venu Govindaraju, University at Buffalo, SUNY Daniel Lopresti, Lehigh University Prem Natarajan, Raytheon BBN Technologies

DAS General Chairs

# Workshop-at-a-Glance

| DAS 2010 Workshop Schedule |  |  |   |  |
|----------------------------|--|--|---|--|
| Time                       | Wednesday: June 9, 2010                                    | Thursday: June 10, 2010                                    | Friday: June 11, 2010   |  |
| 0800                       | Breakfast (PreFunction)                                    | Breakfast (PreAssembly)                                    | Breakfast (PreAssembly)   |  |
|                            | Registration (Pre-Assembly)                                |  |   |  |
| 0900                       | Welcome followed by  |  |   |  |
| 0920                       | Keynote address:   | Oral Session 4   | Oral Session 5  |  |
| 0940                       | Prof. George Nagy Chair: David Doermann                    | Chair: George Thoma<br>(Belvidere Ballroom)                | Chair: Horst Bunke<br>(Belvidere Ballroom)                            |  |
| 1000                       | (Belvidere Ballroom)                                       | (serridere sameem)   | (Derridere Dameern)   |  |
| 1020                       | Morning Break<br>(PreFunction)                             | Morning Break<br>(PreAssembly)                             | Morning Break<br>(PreAssembly)  |  |
| 1040                       | Oral Session 1   | Short paper Oral Session                                   |   |  |
| 1100                       | Chair: Koichi Kise   | Chair: Henry Baird   | Oral Session 6 Chair: Elisa Barney Smith                              |  |
| 1120                       | (Belvidere Ballroom)                                       | (Belvidere Ballroom)                                       | (Belvidere Ballroom)  |  |
| 1140                       |  | Working Group set-up                                       | ,   |  |
| 1200                       | Lunch (On your own)*                                       | Lunch (On your own)*                                       | Lunch (On your own)*  |  |
| 1320                       |  | Working Group Discussions                                  | <b>Oral Session 7</b><br>Chair: Thomas Breuel<br>(Belvidere Ballroom) |  |
| 1340                       | Oral Session 2 Chair: Masaki Nakagawa (Belvidere Ballroom) | (Boardroom/Copley/<br>Mariner)                             |   |  |
| 1400                       |  |  |   |  |
| 1420                       |  | Afternoon Break<br>(PreAssembly)                           |   |  |
| 1440                       | Afternoon Break<br>(PreAssembly)                           | Poster / Demo Session 2<br>(Westminster<br>& PreAsssembly) | Afternoon Break<br>(PreAssembly)                                      |  |
| 1500                       |  |  | Working Group Discussions<br>(Boardroom/Copley/                       |  |
| 1520                       | Oral Session 3   |  |   |  |
| 1540                       | Chair: Andreas Dengel (Belvidere Ballroom)                 |  | Mariner)  |  |
| 1600                       |  | Excursion<br>(Boston Duck Tour)                            | Working Group Plenary   |  |
| 1620                       | Reception (Westminster                                     |  | Reports<br>(Belvidere Ballroom)                                       |  |
| 1640                       | & PreAssembly)   |  | Workshop Conclusion   |  |
| 1700                       |  |  |   |  |
| 1720                       | Poster / Demo Session 1 (Westminster                       |  |   |  |
| 1740                       | & PreAssembly)   |  |   |  |
| 1800                       |  | Break  |   |  |
| 1830                       |  | Banquet  |   |  |
|                            |  | (Top of The Hub)   |   |  |

<sup>\*</sup>Options for lunch can be found in your registration packet.

# **Workshop Detailed Schedule**

| Wednesday - June 9, 2010       |  | •  |  |
|--------------------------------|--|--|--|
| 0900 - 1020                    |  | Welcome and Keynote Talk (Belvidere Ballroom)  |  |
|                                |  | Keynote Talk: Document Systems Analysis: Testing, Testing, Testing: Prof. George Nagy, RPI.  |  |
| 1                              | 020 - 1040   | Coffee Break (PreFunction)   |  |
| 1                              | 040 - 1200   | Oral Session 1 - Ground-truth and Datasets 1 (Belvidere Ballroom)  |  |
| _                              | 1040   | Ground Truth Creation for Handwriting Recognition in Historical Documents - Andreas<br>Fischer, Emanuel Indermühle, Horst Bunke, Gabriel Viehhauser, and Michael Stolz   |  |
|                                | 1100   | IBN SINA: A database for research on processing and understanding of Arabic manuscripts images - Reza Farrahi Moghaddam, Mohamed Cheriet, Mathias Adankon, Kostyantyn Filonenko, and Robert Wisnovsky  |  |
| _                              | 1120   | A Framework for the Assessment of Text Extraction Algorithms on Complex Colour Images -<br>Antonio Clavelli, Dimosthenis Karatzas, and Josep Lladós  |  |
|                                | 1140   | An analysis of binarization ground truthing - Elisa H. Barney Smith  |  |
|                                | 200 - 1320   | Lunch (On your own)  |  |
| 1                              | 320 - 1440   | Oral Session 2 - Handwriting Recognition (Belvidere Ballroom)  |  |
| _                              | 1320   | Toward Affine Recognition of Handwritten Mathematical Characters - Oleg Golubitsky,<br>Vadim Mazalov, and Stephen Watt   |  |
|                                | 1340   | Data-Embedding Pen — Augmenting Ink Strokes with Meta-Information - <i>Marcus Liwicki, Seiichi Uchida, Masakazu Iwamura, Shinichiro Omachi, and Koichi Kise</i>  |  |
|                                | 1400   | Gabor Features for Offline Arabic Handwriting Recognition - Jin Chen, Huaigu Cao, Rohit Prasad, Anurag Bhardwaj, and Premkumar Natarajan   |  |
|                                | 1420   | Improved Classification through Runoff Elections - Oleg Golubitsky, and Stephen Watt   |  |
| 1                              | 440 - 1500   | Coffee Break (PreAssembly)   |  |
| 1                              | 500 - 1620   | Oral Session 3 - Layout Analysis 1 (Belvidere Ballroom)  |  |
|                                | 1500   | Table Detection in Heterogeneous Documents - Faisal Shafait, and Ray Smith   |  |
|                                | 1520   | Context-Aware and Content-Based Dynamic Voronoi Page Segmentation - Mudit Agrawal,   |  |
|                                |  | and David Doermann   |  |
| _                              | 1540   | and David Doermann  Analysis and Taxonomy of Column Header Categories for Web Tables - Sharad Seth, Ramana C. Jandhyala, Mukkai Krishnamoorthy, and George Nagy  |  |
| _                              | 1540<br>1600   | Analysis and Taxonomy of Column Header Categories for Web Tables - Sharad Seth, Ra-  |  |
| 1                              |  | Analysis and Taxonomy of Column Header Categories for Web Tables - Sharad Seth, Ramana C. Jandhyala, Mukkai Krishnamoorthy, and George Nagy  Memory-Based Recognition of Camera-Captured Characters - Masakazu Iwamura, Tomo   |  |
|                                | 1600   | Analysis and Taxonomy of Column Header Categories for Web Tables - Sharad Seth, Ramana C. Jandhyala, Mukkai Krishnamoorthy, and George Nagy  Memory-Based Recognition of Camera-Captured Characters - Masakazu Iwamura, Tomohiko Tsuji, and Koichi Kise  |  |
| <b>1</b><br>P1                 | 1600<br>620 - 1700<br>700 - 1830<br>-1: Typefac  | Analysis and Taxonomy of Column Header Categories for Web Tables - Sharad Seth, Ramana C. Jandhyala, Mukkai Krishnamoorthy, and George Nagy  Memory-Based Recognition of Camera-Captured Characters - Masakazu Iwamura, Tomohiko Tsuji, and Koichi Kise  Reception (Westminster & PreAssembly)  Poster/Demo Session 1 (Westminster & PreAssembly)  e Personality Traits and Their Design Characteristics - Ying Li, and Ching Y. Suen  |  |
| P1                             | 1600<br>620 - 1700<br>700 - 1830<br>-1: Typefac  | Analysis and Taxonomy of Column Header Categories for Web Tables - Sharad Seth, Ramana C. Jandhyala, Mukkai Krishnamoorthy, and George Nagy  Memory-Based Recognition of Camera-Captured Characters - Masakazu Iwamura, Tomohiko Tsuji, and Koichi Kise  Reception (Westminster & PreAssembly)  Poster/Demo Session 1 (Westminster & PreAssembly)  |  |
| P1 P1 Ha                       | 1600<br>620 - 1700<br>700 - 1830<br>-1: Typefac<br>-2: Virtual F<br>aris Dindo<br>-3: A Bag of   | Analysis and Taxonomy of Column Header Categories for Web Tables - Sharad Seth, Ramana C. Jandhyala, Mukkai Krishnamoorthy, and George Nagy  Memory-Based Recognition of Camera-Captured Characters - Masakazu Iwamura, Tomohiko Tsuji, and Koichi Kise  Reception (Westminster & PreAssembly)  Poster/Demo Session 1 (Westminster & PreAssembly)  e Personality Traits and Their Design Characteristics - Ying Li, and Ching Y. Suen  |  |
| P1 P1 P1 Ali P1                | 1600<br>620 - 1700<br>700 - 1830<br>1: Typefac<br>2: Virtual F<br>aris Dindo<br>3: A Bag of<br>icia Fornés,<br>4: Use of N   | Analysis and Taxonomy of Column Header Categories for Web Tables - Sharad Seth, Ramana C. Jandhyala, Mukkai Krishnamoorthy, and George Nagy  Memory-Based Recognition of Camera-Captured Characters - Masakazu Iwamura, Tomohiko Tsuji, and Koichi Kise  Reception (Westminster & PreAssembly)  Poster/Demo Session 1 (Westminster & PreAssembly)  e Personality Traits and Their Design Characteristics - Ying Li, and Ching Y. Suen  Restoration of Old Photos for Non-Expert Users - Giuseppe Mazzola, Edoardo Ardizzone, and   |  |
| P1 P1 Ali P1 Ch                | 1600<br>620 - 1700<br>700 - 1830<br>-1: Typefact<br>-2: Virtual F<br>aris Dindo<br>-3: A Bag of<br>icia Fornés,<br>-4: Use of N<br>audhury, M  | Analysis and Taxonomy of Column Header Categories for Web Tables - Sharad Seth, Ramana C. Jandhyala, Mukkai Krishnamoorthy, and George Nagy  Memory-Based Recognition of Camera-Captured Characters - Masakazu Iwamura, Tomohiko Tsuji, and Koichi Kise  Reception (Westminster & PreAssembly)  Poster/Demo Session 1 (Westminster & PreAssembly)  e Personality Traits and Their Design Characteristics - Ying Li, and Ching Y. Suen  Restoration of Old Photos for Non-Expert Users - Giuseppe Mazzola, Edoardo Ardizzone, and  f Notes Approach to Writer Identification in Old Handwritten Music Scores - Albert Gordo, Ernest Valveny, and Josep Llados  MKL as Symbol Classifier for Gujarati Character Recognition - Ehtesham Hassan, Santanu                               |  |
| P1 P1 P1 P1 Ali P1 Ch P1 P1 P1 | 1600<br>620 - 1700<br>700 - 1830<br>-1: Typefac<br>-2: Virtual F<br>aris Dindo<br>-3: A Bag of<br>icia Fornés,<br>-4: Use of N<br>audhury, M<br>-5: Docume<br>. Breuel<br>-6: A Skelet | Analysis and Taxonomy of Column Header Categories for Web Tables - Sharad Seth, Ramana C. Jandhyala, Mukkai Krishnamoorthy, and George Nagy  Memory-Based Recognition of Camera-Captured Characters - Masakazu Iwamura, Tomohiko Tsuji, and Koichi Kise  Reception (Westminster & PreAssembly)  Poster/Demo Session 1 (Westminster & PreAssembly)  e Personality Traits and Their Design Characteristics - Ying Li, and Ching Y. Suen Restoration of Old Photos for Non-Expert Users - Giuseppe Mazzola, Edoardo Ardizzone, and  f Notes Approach to Writer Identification in Old Handwritten Music Scores - Albert Gordo, Ernest Valveny, and Josep Llados  MKL as Symbol Classifier for Gujarati Character Recognition - Ehtesham Hassan, Santanu Il Gopal, and Jignesh Dholakia |  |

P1-8: An Impact of Lingustic Features on Automated Classification of OCR Texts - *Gudila P. Moshi, Lazaro S.P. Busagala, Wataru Ohyama, Tetsushi Wakabayashi, and Fumitaka Kimura* 

mara Palaiahnakote, Anjan Dutt, Chew Lim Tan, and Umapada Pal

P1-7: A New Wavelet-Median-Moment based Method for Mult-Oriented Video Text Detection - Shivaku-

- P1-9: Form Recognition from ink strokes on tablet *De Cao Tran, Patrick Franco, and Jean-Marc Ogier*P1-10: Methodological Considerations on the INEX Structure Extraction Competition *Hervé Déjean, and Jean-Luc Meunier*
- P1-11: Expansion of Queries and Databases for Improving the Retrieval Accuracy of Document Portions Koichi Kise, Megumi Chikano, Kazumasa Iwata, Masakazu Iwamura, Seiichi Uchida, and Shinichiro Omachi
- P1-12: Higher Order MRF for Foreground-Background Separation in Multispectral Images of Historical Manuscripts *Martin Lettner, and Robert Sablatnig*
- P1-13: Introducing a New Image Dissimilarity Measure with an Application to Character Image Clustering in Degraded Historical Documents *Sebastian Colutto*
- P1-14: smartFIX Statistics Towards Systematic Document Analysis Performance Evaluation and Optimization Benjamin Seidler, Markus Ebbecke, and Michael Gillmann
- P1-15: A Polar-based Logo Representation based on Topological and Colour Features Farshad Nourbakhsh, Dimosthenis Karatzas, and Ernest Valveny
- P1-16: A Histogram-based Technique for Automatic Threshold Assessment in a Run Length Smoothing-based Algorithm Stefano Ferilli, Teresa M.A. Basile, and Floriana Esposito
- based Algorithm Stefano Ferilli, Teresa M.A. Basile, and Floriana Esposito
  P1-17: Latent Dirichlet Allocation Based Writer Identification in Offline Handwriting Anurag Bhardwaj,
- Manavender Malgireddy, Srirangaraj Setlur, Venu Govindaraju, and Sitaram Ramachandrula
  P1-18: Towards More Effective Distance Functions for Word Image Matching Raman Jain, C. V. Jawahar
- P1-19: Writer clustering using Multi Scale Integral Orientation Features Asim Imdad Wagan, Stephane Bres and Hubert Emptoz
- P1-20: Connected Component level Multiscript Identification from Ancient Document Images Sheikh Faisal Rashid, Faisal Shafait and Thomas Breuel
- P1-21: Bibliographic information Extraction of Document Title Pages by Combining FDA's N-best Results and Layout DP-Matching Masakazu Fujio, Takeshi Nagasaki and Toshikazu Takahashi
- P1-22: A Visual Perception Approach to Segment Images of Historical Documents Carlos Mello
- P1-23: Improvements in Optical Structure Recognition Application *Igor Filippov, Marc Nicklaus and John Kinney (Demo)*
- P1-24: HIT-OR3C: An Opening Recognition Corpus for Chinese Characters Shusen Zhou, Qingcai Chen, and Xiaolong Wang (*Demo*)
- P1-25: The Data Embedding Pen: Embedding Knowledge in Paper Marcus Liwicki & Seiichi Uchida(Demo)

|             | Thursday - June 10, 2010  |  |
|-------------|---|--|
| 0900 - 1020 | 1020 Oral Session 4 - Ground-truth and Datasets 2 (Belvidere Ballroom)  |  |
| 0900        | IAM-OnDo-database: an Online Handwritten Document Database with Non-uniform Contents - Emanuel Indermühle, Marcus Liwicki, and Horst Bunke  |  |
| 0920        | Document Analysis Issues in Reading Optical Scan Ballots - Daniel Lopresti, George Nagy, and Elisa Barney Smith   |  |
| 0940        | An open approach towards the benchmarking of table structure recognition systems - Asif Shahab, Faisal Shafait, Thomas Kieninger, and Andreas Dengel  |  |
| 1000        | Investigator Name Recognition from Medical Journal Articles: A Comparative Study of SVM and Structural SVM - <i>Xiaoli Zhang, Jie Zou, Daniel X. Le, and George Thoma</i>   |  |
| 1020 - 1040 | Coffee Break (PreAssembly)  |  |
| 1040 - 1140 | Short Paper Oral Session (Belvidere Ballroom)   |  |
| 1040 - 1140 | Short Paper Oral Session (Belvidere Ballroom)   |  |
| 1040        | Table Metadata: Headers, Augmentations and Aggregates - George Nagy, Mukkai Krish-<br>namoorthy, Raghav Padmanabhan, Ramana C. Jandhyala and William Silversmith  |  |
|             | Table Metadata: Headers, Augmentations and Aggregates - George Nagy, Mukkai Krish-  |  |
| 1040        | Table Metadata: Headers, Augmentations and Aggregates - George Nagy, Mukkai Krishnamoorthy, Raghav Padmanabhan, Ramana C. Jandhyala and William Silversmith  Robust Recognition Method of Chemical Structure Images for Japanese Published Patent   |  |
| 1040        | Table Metadata: Headers, Augmentations and Aggregates - George Nagy, Mukkai Krishnamoorthy, Raghav Padmanabhan, Ramana C. Jandhyala and William Silversmith  Robust Recognition Method of Chemical Structure Images for Japanese Published Patent Applications - Akio Fujiyoshi, Koji Nakagawa and Masakazu Suzuki  DynaQ - Faceted Search for Document Retrieval - Christian Reuschling, Stefan Agne and |  |

Working Group Set up (Belvidere Ballroom)

Lunch (On your own)

1140 - 1200

1200 - 1320

| 1320 - 1420   | Working Group Discussions |  |
|---|---------------------------|--|
| 1420 - 1540   | Coffee Break              |  |
| 1540 - 1600   | Poster/Demo Session 2     |  |
| P2-1: Improving XED for extracting content from Arabic PDFs - Karim Hadjar, Rolf Ingold |                           |  |

- P2-2: A Kernel-based Approach to Document Retrieval Albert Gordo, Jaume Gibert, Ernest Valveny, Marçal Rusiñol
- P2-3: Associating Figures with Descriptions for Patent Documents Linlin Li, Chew Lim Tan
- P2-4: Document Analysis Applied to Fragments: Feature Set for the Reconstruction of Torn Documents -Markus Diem, Florian Kleber, Robert Sablatnia
- P2-5: Page Frame Detection for Double Page Document Images Nikolaos Stamatopoulos, Basilis Gatos, Thodoris Georgiou
- P2-6: Automatic Unsupervised Parameter Selection for Character Segmentation Giorgos Vamvakas, Nikolaos Stamatopoulos, Basilis Gatos, Stavros Perantonis
- P2-7: Word Spotting in Alice's Adventures Underground using Multi Scale Integral Orientation Features -Asim Wagan, Stephane Bres, Hubert Emptoz
- P2-8: A New approach for centerline extraction in handwritten strokes An Application to the constitution of a code book - Hani Daher, Veronique Eglin, Stephane Bres, Nicole Vincent
- P2-9: Safely Selecting Subsets of Training Data Dawei Yin, Chang An, Henry Baird
- P2-10: Improving Handwriting Recognition by the Use of Semantic Information Marcus Liwicki, Hassan Mohamed Abou Eisha, Andreas Dengel
- P2-11: The BBN Document Analysis Service: A Platform for Multilingual Document Translation Ehry Mac-Rostie, Rohit Prasad, Stephen Rawls, Matin Kamali, Huaigu Cao, Premkumar Natarajan
- P2-12: Ground-Truthed Dataset of Chemical Structure Images in Japanese Published Patent Applications -Koji Nakagawa, Akio Fujiyoshi, Masakazu Suzuki
- P2-13: Techniques for Static Handwriting Trajectory Recovery: A Survey Vu Nguyen, Michael Blumenstein
- P2-14: Detecting and Recognizing Tables in Spreadsheets Iyad Abu Doush, Enrico Pontelli
- P2-15: Touch & Write A Multi-Touch Table with Pen-Input Marcus Liwicki, Saher Mohamed El-Neklawy, Andreas Dengel
- P2-16: Faithful Mathematical Formula Recognition from PDF Documents Josef Baker, Alan Sexton, Volker Sorge
- P2-17: A Post-Processing Scheme for Malayalam using Statistical Sub-character Language Models -Karthika Mohan, C.V. Jawahar
- P2-18: An Eigen Value Based Approach for Text Detection in Video Guru D. S, Manjunath S, Shivakumara Palaiahnakote, Chew-Lim Tan
- P2-19: A Branch and Bound Algorithm for Graphical Symbol Recognition in Document Images Nibal Nayef and Thomas M. Breuel
- P2-20: Metadata for Structured Document Datasets Henry F. Korth, Dezhao Song and Jeff Heflin
- P2-21: Elastic matching in linear time and constant space Scott MacLean and George Labahn
- P2-22: Performance Evaluation of Curled Textlines Segmentation Algorithms Syed Saqib Bukhari, Faisal Shafait and Thomas M. Breuel
- P2-23: Using Audio Based Disambiguation for Improving Handwritten Mathematical Content Recognition in Classroom Videos - Smita Vemulapalli and Monson H. Hayes III
- P2-24: A Camera-Based Information Acquisition Interface Masakazu Iwamura, Koichi Kise(Demo)
- P2-25: The DAE (Document Analysis and Exploitation) Project Bart Lamiroy (Demo)
- P2-26: GEDI A Groundtruthing Environment for Document Images David Doermann and Elena Zotkina (Demo)
- P2-27:DynaQ Faceted Search for Document Retrieval Christian Reuschling, Stefan Agne and Andreas Dengel (Demo)

| 1600 - 1800 | Excursion - Boston Duck Tour        |
|-------------|-------------------------------------|
| 1800 - 1830 | Break                               |
| 1830 -      | Banquet - Top of the Hub Restaurant |

<sup>\*</sup>Instructions regarding logistics for the Excursion and Banquet can be found in your registration packet.

|             | Friday - June 11, 2010  |  |
|-------------|---|--|
| 0900 - 1020 | Oral Session 5 - Text Processing (Belvidere Ballroom)   |  |
| 0900        | Overlapped Text Segmentation Using Markov Random Field and Aggregation - <i>Xujun Peng, Srirangaraj Setlur, Venu Govindaraju, and Ramachandrula Sitaram</i>     |  |
| 0920        | Handwritten Arabic Text Line Segmentation using Affinity Propagation - Jayant Kumar, Wa<br>Abd-Almageed, Le Kang, and David Doermann                            |  |
| 0940        | Text Extraction From Graphical Document Images Using Sparse Representation - <i>Thai V. Hoang, and Salvatore Tabbone</i>  |  |
| 1000        | Occluded text restoration and recognition - Lanlan Chang, Jun Sun, Misako Suwa, Hiroaki<br>Takebe, Yuan He, and Satoshi Naoi                                    |  |
| 1020 - 1040 | Coffee Break (PreAssembly)  |  |
| 1040 - 1200 | Oral Session 6 - Layout Analysis 2 (Belvidere Ballroom)   |  |
| 1040        | Binarization of Historical Document Images Using the Local Maximum and Minimum - Su Bolan, Lu Shijian, and Chew Lim Tan   |  |
| 1100        | A System to Detect Rooms in Architectural Floor Plan Images - Sébastien Macé, Hervé Locteau, Ernest Valveny, and Salvatore Tabbone                              |  |
| 1120        | Information extraction by finding repeated structure - Evgeniy Bart, and Prateek Sarkar   |  |
| 1140        | Document Image Segmentation using Discriminative Learning over Connected Components  - Syed Saqib Bukhari, Mayce Ibrahim Ali, Faisal Shafait, and Thomas Breuel |  |
| 1200 - 1320 | Lunch (On your own)   |  |
| 1320 - 1440 | Oral Session 7 - Recognition and Retrieval (Belvidere Ballroom)   |  |
| 1320        | Query Driven Word Retrieval in Graphical Documents - Partha Roy, Umapada Pal, and Josep Llados  |  |
| 1340        | Analysis of Whole-Book Recognition - Pingping Xiu, and Henry Baird  |  |
| 1400        | Nearest Neighbor based Collection OCR - Pramod Kompalli, Jawahar C. V., & Manmatha R.   |  |
| 1420        | Efficient Logo Retrieval Through Hashing Shape Context Descriptors - Marçal Rusiñol, and Josep Lladós   |  |
| 1440 - 1500 | Coffee Break (PreAssembly)  |  |
| 1500 - 1600 | Working Group Discussions (Boardroom/Copley/Mariner)  |  |
| 1600 - 1640 | Working Group Plenary Reports (Belvidere Ballroom)  |  |

# Belvidere Ballroom Pre-Function Westminster Pre-Assembly South Condor Maverick A Moverick B Bevator Lobby Maverick B Bevator Cobby Condor Copley Boardroom Mariner Copley

1640

#### Hilton Back Bay Floor Plan—2nd Floor

#### Excursion—Boston Duck Tours

**Workshop Ends - Closing Remarks** 

(www.bostonducktours.com): The fun begins as soon as you board your "DUCK", a W.W.II style amphibious landing vehicle. You'll cruise by all the places that make Boston the birthplace of freedom and a city of firsts. Your ConDUCKtor will be giving you lots of little known facts and interesting insights about the unique and wonderful city that is Boston. A "Splashdown" finale into the Charles River will give you a breathtaking view of the Boston and Cambridge skylines.

#### Banquet location - Top of the Hub

(www.topofthehub.net): One of Boston's finest dining destinations, the award-winning Top of the Hub features outstanding cuisine, impeccable service, and a sophisticated ambience. Soaring 52 floors above the Back Bay, the restaurant and adjoining Skywalk Observatory offer guests a spectacular view of the Boston skyline from every table.

# **DAS 2010 Organizing Committee**

### **General Chairs**

David Doermann Venu Govindaraju
University of Maryland University at Buffalo

Daniel Lopresti Prem Natarajan
Lehigh University Raytheon BBN Technologies

# **Publications and Web Chair**

Srirangaraj Setlur University at Buffalo

# **Finance and Local Arrangements Committee**

# David Frampton Rohit Prasad (Chair) Laura Stephens

**Raytheon BBN Technologies** 

# **Program Committee**

Gady Agam (USA) Adel M. Alimi (Tunisia) Apostolos Antonacopoulos (UK) Henry Baird (USA) Elisa Barney Smith (USA) Abdel Belaïd (France) Kathrin Berkner (USA) Thomas Breuel (Germany) Horst Bunke (Switzerland) Mohamed Cheriet (Canada) Andreas Dengel (Germany) Xiaoqing Ding (China) Hiromichi Fujisawa (Japan) Tin Kam Ho (USA) Jianying Hu (USA) Rolf Ingold (Switzerland) Masakazu Iwamura (Japan) Dimosthenis Karatzas (Spain) Soo-Hyung Kim (Korea)

Koichi Kise (Japan)

Rafael Dueire Lins (Brazil) Cheng-Lin Liu (China) Marcus Liwicki (Germany) R. Manmatha (USA) Simone Marinai (Italy) Masaki Nakagawa (Japan) Satoshi Naoi (Japan) II-Seok Oh (Korea) Shinichiro Omachi (Japan) Umapada Pal (India) Hiroshi Sako (Japan) Venkata Subramaniam (India) Jun Sun (China) Kazem Taghva (USA) Chew Lim Tan (Singapore) George Thoma (USA) Karl Tombre (France) Seiichi Uchida (Japan) Berin Yanikoglu (Turkey)

Laurence Likforman-Sulem (France)