Document Recognition and Retrieval XVIII (EI113)

Part of the IS&T/SPIE 23rd Annual Symposium on Electronic Imaging 2011 23-27 January 2011 • San Francisco, CA USA

Conference Chairs: **Gady Agam**, Illinois Institute of Technology; **Christian Viard-Gaudin**, Univ. of Nantes (France);

Program Committee: **Apostolos Antonacopoulos**, Univ. of Salford (United Kingdom); **Elisa H. Barney Smith**, Boise State Univ.; **Kathrin Berkner**, Ricoh Innovations, Inc.; **Xiaoqing Ding**, Tsinghua Univ. (China); **David S. Doermann**, Univ. of Maryland/College Park; **Oleg Golubitsky**, Univ. of Western Ontario (Canada); **Jianying Hu**, IBM Thomas J. Watson Research Ctr.; **Laurence Likforman-Sulem**, Telecom ParisTech (France); **Marcus Liwicki**, DFKI (Germany); **Xiaofan Lin**, Vobile Inc.; **Daniel P. Lopresti**, Lehigh Univ.; **Hiroshi Sako**, Hitachi, Ltd. (Japan); **Lambert R. B. Schomaker**, Univ. of Groningen (Netherlands); **Sargur N. Srihari**, Univ. at Buffalo; **Venkata Subramaniam**, IBM India Research Lab. (India); **Kazem Taghva**, Univ. of Nevada/Las Vegas; **George R. Thoma**, National Library of Medicine; **Alessandro Vinciarelli**, University of Glasgow (United Kingdom); **Berrin Yanikoglu**, Sabanci Univ. (Turkey); **Jie Zou**, National Library of Medicine

We are pleased to announce the 18th Document Recognition and Retrieval Conference (DRR), to be held on 23-27 January 2011, in San Francisco, CA, USA. DRR is an international conference focused on state-of-the-art research in document recognition and retrieval, for offline, online and web documents. The conference is part of the *Electronic Imaging Symposium*, which brings together researchers from various backgrounds related to electronic imaging for an exciting research event. The conference will include oral/poster presentations, invited talks and invited papers. Accepted papers will be published in DRR Proceedings. For the fifth year, the **Best Student Paper** will be selected among papers whose first authors are full-time students. Note that after many years in San Jose, the conference is moving to San Francisco. Additional details and updated information of this conference can be found at http://www.cs.iit.edu/~drr2011

Recognizing handwritten or degraded machine printed documents (e.g. faxed and old/historical documents) remains a challenging problem. Beyond OCR, document recognition includes the recovery of a document's logical structure and format. With successful layout analysis and recognition, document recognition aims to fully reconstruct a document in electronic form, in its original format (fonts, layout etc.). Among the challenges for machine-printed documents are complex layouts (text written on images, complex backgrounds, etc.), degraded and noisy documents, and robust recognition of tables and equations. Handwritten documents with unconstrained writing style pose additional challenges due to increased variability and segmentation ambiguities. Handwritten documents can be processed both online (where temporal stroke information is available) and offline. Non-textual elements in documents form another class of interesting problems. These include the extraction and recognition logos and signatures, and the conversion of line drawings in documents from raster to vector format, thus creating graphical objects endowed with semantic meaning. Web documents pose both similar and new challenges. We are soliciting papers describing algorithms and systems in all aspects of document recognition and retrieval, for offline, online and Web documents.

One of the primary reasons for digitizing existing paper materials is to simplify retrieval and organization of information. In this regard we are particularly interested in papers which address any of the following issues: retrieval in the presence of noise; retrieval based on sketches, images, tables, diagrams or other non-linguistic objects that appear in the document; retrieval based on text appearing with non-standard alignment, in images or graphics; recognition and tagging of mathematical arrays and equations which serve as indicators of subject content or methodology used in the document; novel methods for retrieval and organization of information based on text or other information in a document. Papers addressing retrieval-specific issues are encouraged to use standard performance metrics such as ROC and precision-recall curves.

Papers are solicited in, but not limited to, the following areas:

Document Recognition
- document segmentation and layout analysis
- machine-printed and handwritten text recognition
- identification and recognition of tables or equations
- processing of degraded (e.g. faxed) or historical documents
- processing of multilingual documents
- filtering, enhancement, and compression techniques for document images
- performance metrics
- document degradation models
- web document analysis (including wikis and blogs)
- video-, camera-, and mobile phone-based OCR
- recognition of text in natural scene images
- graphics recognition (line-art, maps, and technical drawings)
- symbol, signature, and logo recognition
- document style recognition, writer identification
- document analysis and synthesis for digital publishing (template reuse and layout generation for new contents)
- system engineering, systems, and quality assurance methods towards large-scale digital libraries
- information extraction from forms
- document analysis techniques for electronic voting systems

Document Retrieval
- keyword spotting in document images
- approximate string matching algorithms for OCR'ed text
- summarization of text documents and imaged documents
- text categorization from imaged documents
- entity tagging using OCR'ed text
- retrieval of noisy text documents (messages, blogs, etc.)
- nontextual retrieval (e.g. using graphics and images)
- recovery and use of logical structure for retrieval
- cross-language and multi-lingual retrieval
- benchmarking and evaluation issues
- relevance feedback techniques for document retrieval
- impact of recognition accuracy on retrieval effectiveness

Note: Submissions to Document Recognition and Retrieval XVIII should contain abbreviated papers (5-7 pages). Submissions should be informative, describe the problem that is addressed by the paper, the original contribution in the paper, the way it relates to existing work, and provide experimental/theoretical evaluation. Final manuscripts to be published in the proceedings are expected to be 8-12 pages long. Questions concerning the conference could be addressed to: drr2011@cs.iit.edu

Important dates:
28 June 2010:   Abbreviated papers due
9 August 2010:   Acceptance notice
15 November 2010: Final manuscripts due