

ANURAG BHARDWAJ

St. 202, 520 Lee Entrance,
Amherst, NY 14228
<http://www.cse.buffalo.edu/~ab94>

Phone: (716) 645-6164
Email: ab94@buffalo.edu
Alt: ab94@cedar.buffalo.edu

PARTICULARS

EDUCATION

University at Buffalo, The State University of New York
Ph. D. in Computer Science and Engineering
Amherst, NY
Sep 2008 - Sep, 2010(Expected)

University at Buffalo, The State University of New York
M. S. in Computer Science and Engineering
Amherst, NY
Aug 2006 - Sep 2008

National Institute of Technology
B. Tech. in Computer Engineering
Kurukshetra, India
Aug 2001 - May 2005

CURRENT STATUS

U.S. Student Visa, Citizen of India.

RESEARCH INTERESTS

My research interests are machine learning, natural language processing and document image analysis. I like to apply probabilistic learning algorithms for document image recognition and retrieval. My other areas of interest are biometrics, smart environments and human-computer interaction.

DISSERTATION

Title: "Efficient Digital Access to Handwritten Documents"
Advisor: Prof. Venu Govindaraju

My thesis focuses on developing a unified framework for handwritten document search which combines both text and image attributes for robust ranking and retrieval of documents.

ACADEMIC HONORS

- Graduate Teaching Assistantship, University at Buffalo, Spring 2010.
- Graduate Research Assistantship, University at Buffalo, Spring 2007 - Fall 2009.
- Microsoft Student Ambassador, Asia-Pacific, 2005.
- Merit Scholarship, National Institute of Technology, Kurukshetra, 2003.

RESEARCH EXPERIENCE

- **Research Assistant, Center for Unified Biometrics and Sensors, University at Buffalo**, Jan 2007 - Dec 2009.

Handwritten Document Retrieval, (Sep 2009 - Present): Currently working on combination of rankers based on text and image based features for robust handwritten document retrieval.

Writer Identification and Verification, (Jan 2009 - Aug 2009): Worked on a generative model in form of Latent Dirichlet Allocation(LDA) that automatically infers writing styles from handwritten document collection and uses this information to represent each writer as a distribution over multiple writing style for classifying any unknown writer sample.

Keyword Spotting of Handwritten Documents, (Jan 2007 - Dec 2007): Worked on a novel word segmentation algorithm which is used along with the word recognition likelihood to improve keyword spotting in low-quality handwritten documents.

Recognition and retrieval of Indic Documents, (Jan 2007 - Dec 2007): Developed techniques for script-specific as well script-independent keyword spotting in Indic documents.

- **Research Assistant, Center of Excellence for Document Analysis and Recognition, University at Buffalo**, Oct 2006 - Jan 2007.

Lexeme features for handwriting analysis: Worked on extraction of lexeme features from bigrams such as "th" for the task of writer verification in handwriting.

- **Image Processing Scientist Intern, BBN Technologies, Cambridge, MA**, May - Aug 2009 and June - Aug 2008.

Handwritten Arabic Document Recognition: Worked on stochastic segment modelling for offline handwritten Arabic document recognition.

- **Research Intern, KYOS Systems Inc., Burlington, MA**, May - Aug 2007.

Handwritten Medical Forms Analysis: Developed a date recognizer and a keyword spotting system for handwritten medical forms. Also worked on segmentation of handwritten and machine-print text in the form.

TEACHING EXPERIENCE

- **Teaching Assistant.** *Introduction to Computer Vision and Pattern Recognition*, Spring 2010, Deptt. of Computer Science and Engineering, University at Buffalo.
- **Guest Lecturer.** Seminar on *Digital Libraries*, Tutorial on Feature Extraction and Handwriting Recognition. Spring 2009, Deptt. of Computer Science and Engineering, University at Buffalo.
- **Guest Lecturer.** Seminar on *Markov Models for Pattern Recognition*, Tutorial on HMM Toolkit. Spring 2008, Deptt. of Computer Science and Engineering, University at Buffalo.

WORK EXPERIENCE

- **Software Engineer, IBM Systems and Technology Lab, Bangalore, India** July 2005 - July 2006. Functional Verification of AIX Virtual IO Server and AIX Storage Device Drivers.

PUBLICATIONS

JOURNAL PAPERS

1. Robert Milewski, Venu Govindaraju and Anurag Bhardwaj, "Automatic Recognition of Handwritten Medical Forms for Search Engines", *International Journal on Document Analysis and Recognition*, 11(4), pp. 203–218, 2009.
2. Huaigu Cao, Anurag Bhardwaj and Venu Govindaraju, "A Probabilistic method for Keyword Retrieval in Handwritten Document Images", *Pattern Recognition*, 42(11), pp. 3374–3382, 2009.
3. Faisal Farooq, Anurag Bhardwaj and Venu Govindaraju, "Using Topic Models for OCR Correction", *International Journal on Document Analysis and Recognition*, 12(3), pp. 153–164, 2009.

BOOK CHAPTERS

4. Anurag Bhardwaj, Srirangaraj Setlur and Venu Govindaraju, "Keyword Spotting Techniques for Sanskrit Documents", *Sanskrit Computational Linguistics*, Lecture Notes in Computer Science, pp. 403–416, Vol. 5402, Springer, 2009.
5. Anurag Bhardwaj, Srirangaraj Setlur and Venu Govindaraju, "Word Spotting for Indic Scripts to Facilitate Retrieval", *Guide to OCR for Indic Scripts*, Advances in Pattern Recognition, pp. 285–299, Springer, 2010.

CONFERENCE AND WORKSHOP PAPERS

6. Anurag Bhardwaj, Manavender Malgireddy, Srirangaraj Setlur, Venu Govindaraju and Ramachandrupa Sitaram, "Writer Identification in Offline Handwriting using Topic Models", *Proceedings of NIPS Workshop on Topic Models: Text and Beyond*, Canada, 2009.
7. Venu Govindaraju, Huaigu Cao and Anurag Bhardwaj, "Handwritten Document Retrieval Strategies", *Proceedings of ICDAR Workshop on Noisy Text Analytics (AND)*, Spain, 2009.
8. Prem Natarajan, Krishna Subramanian, Anurag Bhardwaj and Rohit Prasad, "Stochastic Segment Modeling for Offline Handwriting Recognition", *Proceedings of X International Conference on Document Analysis and Recognition (ICDAR)*, Spain, July. 2009.
9. Anurag Bhardwaj, Huaigu Cao and Venu Govindaraju, "Script Identification for Handwritten Word Images", In *Proceedings of XVI Document Recognition and Retrieval*, San Jose, Jan. 2009.
10. Anurag Bhardwaj, Faisal Farooq, Huaigu Cao and Venu Govindaraju, "Topic Based Language Models for OCR Correction", *Proceedings of SIGIR workshop on Noisy Text Analytics (AND)*, Singapore, 2008.
11. Huaigu Cao, Anurag Bhardwaj and Venu Govindaraju, "A Probabilistic method for Keyword Retrieval in Handwritten Document Images", *International Conference on Frontiers in Handwriting Recognition*, 2008.
12. Anurag Bhardwaj, S. Kompalli, S. Setlur, Venu Govindaraju, "An OCR Based Approach for Word Spotting in Devanagari Documents", *Proceedings of XV Document Recognition and Retrieval*, San Jose, Jan. 2008.
13. Damien Jose, Anurag Bhardwaj, Venu Govindaraju, "Transcript Mapping for Handwritten English Documents", *Proceedings of XV Document Recognition and Retrieval*, San Jose, Jan. 2008.
14. Anurag Bhardwaj, Damien Jose and Venu Govindaraju, "Script Independent Word Spotting for Multilingual Documents", *Proceedings of Workshop on Cross Lingual Information Access - IJCNLP*, Hyderabad, Jan. 2008.
15. Anurag Bhardwaj, A. Singh, H. Srinivasan, S. Srihari, "On Lexeme Features of Handwritten Documents", *Proceedings of IX International Conference on Document Analysis and Recognition (ICDAR)*, Brazil, Sept. 2007.
16. Anurag Bhardwaj, Manavender Malgireddy, Srirangaraj Setlur, Venu Govindaraju and Ramachandrupa Sitaram, "Latent Dirichlet Allocation Based Writer Identification in Offline Handwriting", *Document Analysis Systems*, 2010.
17. Jin Chen, Huaigu Cao, Rohit Prasad, Anurag Bhardwaj and Premkumar Natarajan, "Gabor Features for Offline Arabic Handwriting Recognition", *Document Analysis Systems*, 2010.

PAPERS UNDER REVIEW

18. Huaigu Cao, Venu Govindaraju and Anurag Bhardwaj, "Handwritten Document Retrieval Strategies", *International Journal on Document Analysis and Recognition*.
19. Rohit Prasad, Anurag Bhardwaj, Krishna Subramanian, Huaigu Cao and Premkumar Natarajan, "Stochastic Segment Model Adaptation for Offline Handwriting Recognition", *International Conference on Pattern Recognition*, 2010.

PAPERS IN PREPARATION

20. Anurag Bhardwaj and Venu Govindaraju, "Aggregating Text and Image Features for Handwritten Document Search".

TECHNICAL DISCLOSURES

1. Anurag Bhardwaj and Sunil Sagar, "A Parametric Method for Estimation of Software Reliability at High Stress", Technical Disclosure, IBM India Pvt. Ltd., June 2007.

TALKS AND PRESENTATIONS

CONFERENCE TALKS

1. "An OCR Based Approach for Word Spotting in Devanagari Documents", *XV SPIE Document Recognition and Retrieval*, San Jose, Jan. 2008.
2. "Transcript Mapping for Handwritten English Documents", *XV SPIE Document Recognition and Retrieval*, San Jose, Jan. 2008.

3. "A Probabilistic method for Keyword Retrieval in Handwritten Document Images", *Computer Science and Engineering Graduate Research Conference (CSEGSA)*, University at Buffalo, 2008.

POSTER PRESENTATIONS

4. "Writer Identification in Offline Handwriting using Topic Models", *NIPS workshop on Topic Models: Text and Beyond*, Canada, 2009.

PROFESSIONAL ACTIVITIES

- Student Member, IEEE 2008 - Present.
- Student Member, ACM 2008 - Present.

SERVICE

- Reviewer (Journals) - *IEEE Transactions on Systems, Man and Cybernetics - Part B*.
- Reviewer (Conferences) *8th International Workshop on Document Analysis Systems, Nara, Japan, September 16 - 19, 2008*.
- Member, Scientific Committee (Conferences) - *14th International Conference of the International Graphonomics Society Dijon, France, September 13 - 16, 2009*.

LANGUAGES

Proficient in English and Hindi.