Ritendra Datta

Contact Information	Collaborative Innovation Center 4720 Forbes Avenue, Pittsburgh, PA 15213	*Cell Phone:* (814) 441-1851 *E-mail:* ritendra [at] google [dot] com *URL:* http://ritendra.weebly.com

Current

Google Inc., Pittsburgh, PA USA   **May 2009 - present**
Engineer - Search properties - algorithm design, machine learning, system development.

Education

The Pennsylvania State University, University Park, PA USA

Ph.D., Computer Science and Engineering   **August 2004 - February 2009**
- *Advisors:* Dr. James Z. Wang and Dr. Jia Li
- *Cumulative GPA:* 3.85/4.00, *Major GPA:* 3.90/4.00 (CSE/IST courses)
- *Candidacy:* January 2005, *Proposal:* March 2007, *Final Defense:* February 2009.
- *Title:* "Semantics and Aesthetics Inference for Image Search: Statistical Learning Approaches"

Bengal Engineering and Science University, Shibpur, India

Bachelor of Engineering (Honors), Information Technology   **July 2000 - July 2004**
- *Advisor:* Dr. Pabitra K. Ray

Industry Experience

Google Inc., Pittsburgh, PA USA
Summer Intern, Product Search research group   **Summer 2008**
- *Mentor:* Dr. Charles Schafer
- *Project:* Learning-based high precision/recall entity matching for Product Search.

Xerox PARC, Palo Alto, CA USA
Summer Intern, Computing Science Laboratory   **Summer 2007**
- *Mentor:* Dr. Marshall Bern
- *Project:* Learning-based *de novo* sequencing of proteins.

IBM T.J. Watson Research Center, Yorktown Heights, NY USA
Summer Intern, Mathematical Sciences Department   **Summer 2006**
- *Mentors:* Dr. Jianying Hu and Dr. Bonnie Ray
- *Project:* Duration HMM based clustering algorithms for business analytics.

Academic Experience

The Pennsylvania State University, University Park, PA USA
Instructor, IST 210 Bootcamp (full responsibility, 80 students)   **Spring 2006, Spring 2007**
Research Assistant, 3COM Intelligent Information Systems Lab   **Spring 2005 - now**
Teaching Assistant, Digital Image Processing (CSE/EE 485)   **Fall 2004**

Awards and Achievements

Publications have received 660+ citations according to Google Scholar (as of January 2010).

Erdös Number 3, as of March 2008 (P. Erdös $\Rightarrow$ {B. Aronov, F.F. Yao, ...} $\Rightarrow$ M. Bern $\Rightarrow$ Me).

Named by IBM Research as one of eight "Emerging Leaders in Multimedia" for the year 2007.

IBM Student Travel Award, ACM Multimedia, Augsburg, Germany, 2007.

Nominated for Microsoft Ph.D. Fellowship, by Penn State CSE Department, 2006-07.

Paper received 2nd highest peer-rating, ACM Multimedia (of 128 Content Track submissions), 2006.

Glenn E. Singley Memorial Graduate Fellowship in Engineering, Penn State, 2004-05.

University Fellowship for Ph.D. study, Computer Science, SUNY at Stony Brook, 2004-05 (declined).

Ranked 6th in ACM International Collegiate Programming Contest (ICPC), Asia Regional, 2004.

Ranked 56th out of 2000+ participants in Bitwise 2k4 Programming Contest, IIT Kharagpur, 2004.

Qualified MENSA, the international high-IQ society (IQ: 156 in Cattel Scale, 99th percentile), 2001.

Government Scholarship (Best Applied Project), Eastern India Science and Engineering Fair, 2003.

Ranked in top 0.5% (50,000+ applicants) in the WB-JEE, state engineering admission exam, 2000.

Soumitra Biswas Memorial Medal for Best Computer Science Student (high school level), 1997-98.

Program Committee Member, ACM Multimedia 2010.

Program Committee Member, ICME 2010.

Program Committee Member, WWW 2010.

Local Arrangement Chair, ACM MIR Conference, 2010

Program Committee Member, ACM MIR Conference, 2010.

Program Committee Member, LS-MMRM Workshop, ACM Multimedia 2009.

Judge, Penn State Graduate Exhibition, 2009.

Judge, Penn State Graduate Exhibition, 2008.

"Predicting Object Identity using an Ensemble of Predictors," US Patent filed, Sept. 2009.

"Automated Image Annotation based upon Meta-learning over Time," US Patent filed, #12/234159, Sept. 2008.

"Studying Aesthetics in Photographic Images Using a Computational Approach," US Patent filed, #12/116578, May 2008.

"Image-based CAPTCHA Generation System," US Patent filed, #11/668853, January 2007.

Refereed Journals Articles:

23. Ritendra Datta, Jia Li, and James Z. Wang, "Exploiting the Human-Machine Gap in Image Recognition for Designing CAPTCHAs," **IEEE Trans. Information Forensics and Security**, vol. 4, no. 3, pp. 504-518, 2009.

22. Ritendra Datta and Marshall Bern, "Spectrum Fusion: Using Multiple Mass Spectra for De Novo Peptide Sequencing," **Journal of Computational Biology**, vol. 16, no. 8, pp. 1169-1182, 2009.

21. Ritendra Datta, Dhiraj Joshi, Jia Li, and James Z. Wang, "Image Retrieval: Ideas, Influences, and Trends of the New Age," **ACM Computing Surveys**, vol. 40, no. 2, 2008.

20. Ritendra Datta, Weina Ge, Jia Li, and James Z. Wang, "Toward Bridging the Annotation-Retrieval Gap in Image Search," **IEEE Multimedia**, vol. 14, no. 3, pp. 24–35, 2007.

Refereed Conference/Workshop Papers:

19. Neela Sawant, Ritendra Datta, Jia Li, and James Z. Wang, "Quest for Relevant Tags using Local Interaction Networks and Visual Content," **Proceedings of the ACM Multimedia Information Retrieval Conference**, Philadelphia, PA, 2010.

18. Ritendra Datta and James Z. Wang, "ACQUINE: Aesthetic Quality Inference Engine - Real-time Automatic Rating of Photo Aesthetics," **Proceedings of the ACM Multimedia Information Retrieval Conference**, Demonstrations, Philadelphia, PA, 2010.

17. Ritendra Datta, Jianying Hu, and Bonnie Ray, "On Efficient Viterbi Decoding for Hidden semi-Markov Models," **Proceedings of the IEEE International Conference on Pattern Recognition**, (Oral), Tampa, FL, 2008. *(Acceptance rate: 18%)*

16. Ritendra Datta, Jia Li, and James Z. Wang, "Algorithmic Inferencing of Aesthetics and Emotion in Natural Images: An Exposition," **Proceedings of the IEEE International Conference on Image Processing**, San Diego, CA, 2008.

15. Ritendra Datta and Marshall Bern, "Spectrum Fusion: Using Multiple Mass Spectra for De Novo Peptide Sequencing," **Proceedings of the Conference on Research in Computational Molecular Biology (RECOMB)**, 13 pages, Singapore, April 2008. *(Acceptance rate: 18%)*

14. Amitayu Das, Ritendra Datta, Anand Sivasubramaniam, and Bhuvan Urgaonkar, "Predicting Web Cache Behavior Using Stochastic State-Space Models," **Proceedings of the International**

Workshop on Scalable Data Management Applications and Systems, Las Vegas, NV, 2008.

13. <u>Ritendra Datta</u>, Dhiraj Joshi, Jia Li, and James Z. Wang, "Tagging over Time: Real-world Image Annotation by Lightweight Meta-learning," **Proceedings of the ACM Multimedia Conference**, 10 pages, Augsburg, Germany, September 2007. *(Acceptance rate: 17%)*

12. <u>Ritendra Datta</u>, Jia Li, and James Z. Wang, "Learning the Consensus on Visual Quality for Next-Generation Image Management," **Proceedings of the ACM Multimedia Conference**, 4 pages, Augsburg, Germany, September 2007. *(Acceptance rate: 27%)*

11. <u>Ritendra Datta</u>, Jianying Hu, and Bonnie Ray, "Sequence Mining for Business Analytics: Building Project Taxonomies for Resource Demand Forecasting," **Proceedings of the Workshop on Data Mining for Business, PAKDD**, Nanjing, May 2007.

10. <u>Ritendra Datta</u>, Weina Ge, Jia Li and James Z. Wang, "Toward Bridging the Annotation-Retrieval Gap in Image Search by a Generative Modeling Approach," **Proceedings of the ACM Multimedia Conference**, 10 pages, Santa Barbara, CA, October 2006. *(Acceptance rate: 16%)*

9. Dhiraj Joshi, <u>Ritendra Datta</u>, Ziming Zhuang, WP Weiss, Marc Friedenberg, James Z. Wang, and Jia Li, "PARAgrab: A Comprehensive Architecture for Web Image Management and Multimodal Querying," **Proceedings of the International Conference on Very Large Data Bases (VLDB)**, Demonstration, 4 pages, Seoul, Korea, September 2006. *(Acceptance rate: 34%)*

8. <u>Ritendra Datta</u>, Dhiraj Joshi, Jia Li, and James Z. Wang, "Studying Aesthetics in Photographic Images Using a Computational Approach," **Proceedings of the European Conference on Computer Vision (ECCV)**, pp. 288–301, Graz, Austria, May 2006. *(Acceptance rate: 21%)*

7. <u>Ritendra Datta</u>, Jia Li, and James Z. Wang, "IMAGINATION: A Robust Image-based CAPTCHA Generation System," **Proceedings of the ACM Multimedia Conference**, pp. 331–334, Singapore, November 2005. *(Acceptance rate: 29%)*

6. <u>Ritendra Datta</u>, Jia Li, and James Z. Wang, "Content-Based Image Retrieval - Approaches and Trends of the New Age," **Proceedings of the 7th International Workshop on Multimedia Information Retrieval, ACM Multimedia**, pp. 253–262, Singapore, November 2005.

5. Ashish Parulekar, <u>Ritendra Datta</u>, Jia Li, and James Z. Wang, "Large-scale Satellite Image Browsing using Automatic Semantic Categorization and Content-based Retrieval," **Proceedings of the Workshop on Semantic Knowledge in Computer Vision, IEEE International Conference on Computer Vision (ICCV)**, 8 pages, Beijing, China, October 2005.

Refereed Book Chapters:

4. <u>Ritendra Datta</u>, Jianying Hu, and Bonnie Ray, "Sequence Mining for Business Analytics: Building Project Taxonomies for Resource Demand Forecasting," **Applications of Data Mining in E-Business and Finance**, IOS Press, vol. 177, pp. 133-141, 2008.

Technical Reports (Non-refereed):

3. <u>Ritendra Datta</u>, Jia Li, and James Z. Wang, "Adapting Automatic Image Annotation via Meta-learning," **Penn State University Technical Report**, CSE 09-005, 2009.

2. <u>Ritendra Datta</u>, Jia Li, and James Z. Wang, "Exploiting the Human-Machine Gap in Image Recognition for Designing CAPTCHAs," **Penn State University Technical Report**, CSE 08-011, 2008.

1. <u>Ritendra Datta</u>, Jia Li, Ashish Parulekar, and James Z. Wang, "Scalable Remotely Sensed Image Mining Using Supervised Learning and Content-based Retrieval," **Penn State University Technical Report**, CSE 06-019, 2006. 11, in submission to journal.

PRESENTATIONS
- "Algorithmic Inferencing of Aesthetics and Emotion in Natural Images: An Exposition", IEEE ICIP, Special Session on Image Aesthetics, Mood and Emotion, San Diego, CA, October 2008. *(Talk)*
- "Tagging over Time: Real-world Image Annotation by Lightweight Meta-learning", Emerging Leaders in Multimedia, IBM T.J. Watson Research Center, Hawthorne, NY, October 2007.

(Invited Talk)

- "Tagging over Time: Real-world Image Annotation by Lightweight Meta-learning", ACM Multimedia, Augsburg, Germany, September 2007. *(Talk)*
- "Learning the Consensus on Visual Quality for Next-Generation Image Management", ACM Multimedia, Augsburg, Germany, September 2007. *(Poster)*
- "Studying Aesthetics in Photographic Images Using a Computational Approach", ECCV, Graz, Austria, 2006. *(Poster)*
- "CBIR and Annotation Techniques for Museum Images," 33rd MCN Conf., Boston, 2005. *(Invited Talk)*
- "Large-scale Satellite Image Browsing Using Automatic Semantic Categorization and Content-based Retrieval, SKCV Workshop, ICCV, Beijing, 2005. *(Talk)*
- "The Automatic Linguistic Indexing of Pictures System,", CVPR, San Diego, CA, 2005. *(Demo)*

FUNDING PROPOSALS

- NSF CT-ISG: *Exploiting the Image Semantic Gap for Robust Authentication*, Mar. 2008.
- NSF RI-Medium: *Modeling Visual Aesthetics based on Content and Context*, Oct. 2007.

REVIEWING

Journals:
ACM Computing Surveys, IEEE Transactions on Pattern Analysis and Machine Intelligence, IEEE Transactions on Image Processing, IEEE Transactions on Knowledge and Data Engineering, IEEE Transactions on Multimedia, IEEE Transactions on Systems, Man, and Cybernetics-B, IEEE Transactions on Geosciences and Remote Sensing, IEEE Communications Letters, IEEE Multimedia, Journal of Machine Learning Research, International Journal of Computer Vision, Pattern Recognition, Pattern Recognition Letters, Journal of Computational and Graphical Statistics, Technometrics, Neurocomputing Journal.

Conferences and Workshops:
WWW 2010, IEEE ICME 2010, ACM Multimedia Workshop 2009, ACM MIR 2010, SIGGRAPH 2009, IEEE IGARSS 2008, ICIP 2008, CIKM Workshop 2008, ACM Multimedia 2006, ACM SIGIR 2005, ICADL 2005, ACM SIGMOD Workshop 2005, IEEE ICCV Workshop 2005, AAAI Workshop 2005, ACM Multimedia Workshop 2005.

RELEVANT COURSEWORK

Applied Mathematics, Numerical Linear Algebra, Optimization Techniques, Discrete Mathematics. Theory of Statistics, Data Mining, Applied Statistics, Probability and Statistics. Computer Vision, Information Retrieval and Search Engines, Intelligent Control, Bioinformatics. Advanced Databases, Computer Networks, Computer Architecture, Programming Languages.

PROGRAMMING

C, C++, Perl, Matlab, Java, CGI/Perl, XHTML, CSS, PHP, MySQL, UNIX shell script, Prolog, SML, Objective-C (iPhone SDK).

OTHER ACTIVITIES

Volunteer, Association for India's Development (non-profit), Pittsburgh Chapter, 2009-current. Conceived and coordinated an office-wide treasure hunt contest, Google Pittsburgh, 2008. Runners-up, Picture Your World Photo Contest (Category: Culture snapshots), Penn State, 2007. Vocalist/Guitarist/Keyboardist, Indian community music groups, Penn State, 2006-current. Delegate, Graduate Student Association (GSA), Penn State, 2005-06. Member and Volunteer, GSA Cafe Film Series, Penn State, 2005-06. Solved the 3×3×3 Rubik's Cube in $\mu = 137$ seconds ($\sigma = 21$) over 30 measured attempts.

REFERENCES

Prof. James Z. Wang (Ph.D. Advisor)
Associate Professor
College of Information Science and Technology
Penn State University, University Park, PA
jwang@ist.psu.edu, (814) 865-7889

Prof. Jia Li (Ph.D. Advisor)
Associate Professor
Department of Statistics

Penn State University, University Park, PA
jiali@stat.psu.edu, (814) 863-3074

Dr. Jianying Hu
Research Staff Member
Mathematical Sciences Department
IBM T.J. Watson Research Center, Yorktown Heights, NY
jyhu@us.ibm.com, (914) 945-1061

Dr. Marshall Bern
Principal Scientist
Computing Science Laboratory
Palo Alto Research Center (Xerox PARC), Palo Alto, CA
bern@parc.com, (650) 812-4443

Dr. Dhiraj Joshi
Research Scientist
Intelligent Systems Group
Kodak Research Labs, Rochester, NY
dhiraj.joshi@kodak.com, (814) 574-6259

Dr. Charles Schafer
Software Engineer
Google Inc., Pittsburgh, PA
cschafer@google.com