Vebjorn Ljosa

Broad Institute of MIT and Harvard
7 Cambridge Center Rm. 6028C
Cambridge, MA 02142

(978) 206-1024
vebjorn@ljosa.com
http://www.ljosa.com/~ljosa/

Personal statement

My research is focused on data mining for microscopy-based high-throughput screens and profiling experiments, in particular algorithms that extracting succinct signatures of the phenotypical state of individual cells and populations of cells based on microscopy images. My strongest expertise lies in management of uncertain data: indexing, mining, and reasoning with uncertain values and probability distributions of measurements of biological systems. In particular, I am a leader in index structures, query algorithms, and distance metrics for arbitrary probability distributions. My PhD thesis work and subsequent research at the Broad Institute have been heavily oriented toward the analysis needs of real screens and other high-throughput experiments, giving me broad experience in microscopy and issues related to sample preparation and handling in relation to image quality and analysis.

Education

Ph.D. in Computer Science, University of California, Santa Barbara, 2002–2007
Advisor: Prof. Ambuj K. Singh
Dissertation title: "Managing probabilistic data: toward data-driven biology"
Developed algorithms for managing uncertainty in biological data and to obtain, search, and mine probabilistic values.

M.S. in Computer Science, Norwegian University of Science and Technology, Trondheim, Norway, 1996–2002. Thesis work at University of California, Santa Barbara.
Thesis title: "Adaptive data replication and consistency"
Investigated adaptive data replication algorithms in large networks and studied how the efficiency of such algorithms is affected by the consistency condition imposed on the data.

Experience

Computational biologist, Broad Institute of MIT and Harvard 2007–present
Developing data-mining and image-analysis algorithms for microscopy-based high-throughput screens and profiling experiments.

Ph.D. student, University of California, Santa Barbara 2002–2007

M.S. student, Norwegian University of Science and Technology 1996–2002

Summer intern, Telenor Nextra, Oslo, Norway Summer 2001
Developed billing software in Common Lisp.

Summer intern, FAST Search & Transfer, Oslo, Norway Summer 2000
Developed search-engine software in C++, Python, and C.

Co-founder, Initio IT-losninger, Trondheim, Norway 1998–2001
Co-founded the company, which had five full-time employees, and served on the board from 1998 to 2000. Wrote user and account management software for an ISP. Planned corporate wide-area network with more than 40 locations. Set up networks and servers for ISPs and corporate customers. Advised customers on network security.

System administrator, Norwegian University of Science and Technology 1996–1999
Managed the central Unix sytem for students, with 20,000 users.

Awards

Dean's fellowship, University of California, Santa Barbara, 2006–2007

Best student paper award, Fifth IEEE International Conference on Data Mining (ICDM), 2005

Peer-reviewed research articles

Tammy Riklin-Raviv, Vebjorn Ljosa, Annie L. Conery, Frederick M. Ausubel, Anne E. Carpenter, Polina Golland, and Carolina Wählby: "Morphology-Guided Graph Search for Untangling Objects: *C. elegans* Analysis," in *Proceedings of the 13th International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI)*, September, 2010. To appear.

Carolina Wählby, Tammy Riklin-Raviv, Vebjorn Ljosa, Annie L. Conery, Polina Golland, Frederick M. Ausubel, and Anne E. Carpenter: "Resolving clustered worms via probabilistic shape models," in *Proceedings of the IEEE International Symposium on Biomedical Imaging (ISBI)*, p. 552–555, April, 2010, doi:10.1109/ISBI.2010.5490286.

Vebjorn Ljosa and Ambuj K. Singh: "Top-k Spatial Joins of Probabilistic Objects," in *Proceedings of the 24th International Conference on Data Engineering (ICDE)*, p. 566–575, April, 2008, doi:10.1109/ICDE.2008.4497465.

Vebjorn Ljosa and Ambuj K. Singh: "APLA: Indexing Arbitrary Probability Distributions," in *Proceedings of the 23rd International Conference on Data Engineering (ICDE)*, p. 946-955, April, 2007, doi:10.1109/ICDE.2007.367940.

Vebjorn Ljosa and Ambuj K. Singh: "Probabilistic Segmentation and Analysis of Horizontal Cells," in *Proceedings of the 2006 IEEE International Conference on Data Mining (ICDM)*, p. 980-985, December, 2006, doi:10.1109/ICDM.2006.129.

Vebjorn Ljosa and Ambuj K. Singh: "Probabilistic segmentation of horizontal cells," in *Proceedings of the 2006 workshop on multiscale biological imaging, data mining & informatics (Bioimage Informatics)*, p. 39–40, September, 2006.

Vebjorn Ljosa, Arnab Bhattacharya, and Ambuj K. Singh: "LB-Index: A Multi-Resolution Index Structure for Images," in *Proceedings of the 22nd International Conference on Data Engineering (ICDE)*, April, 2006, doi:10.1109/ICDE.2006.85.

Vebjorn Ljosa, Arnab Bhattacharya, and Ambuj K. Singh: "Indexing Spatially Sensitive Distance Measures Using Multi-Resolution Lower Bounds," in *Proceedings of the 10th International Conference on Extending Database Technology (EDBT)*, p. 865-883, March, 2006, doi:10.1007/11687238_51.

Arnab Bhattacharya, Vebjorn Ljosa, Jia-Yu Pan, Mark R. Verardo, Hyungjeong Yang, Christos Faloutsos, and Ambuj K. Singh: "ViVo: Visual vocabulary construction for mining biomedical images," in *Proceedings of the Fifth IEEE International Conference on Data Mining (ICDM)*, p. 50–57, November, 2005, doi:10.1109/ICDM.2005.151. Received one of five "best student paper" awards.

Tamer Kahveci, Vebjorn Ljosa, and Ambuj K. Singh: "Speeding up whole-genome alignment by indexing frequency vectors," *Bioinformatics*, 20(13), p. 2122–2134, 2004, doi:10.1093/bioinformatics/bth212.

Review articles

Vebjorn Ljosa and Anne E. Carpenter: "Introduction to the Quantitative Analysis of Two-Dimensional Fluorescence Microscopy Images for Cell-Based Screening," *PLoS*

Computational Biology, 5(12):e10000603, 2009, doi:10.1371/journal.pcbi.1000603. Peer-reviewed.

Arnab Bhattacharya and Vebjorn Ljosa: "Image Management for Biological Data," in Encyclopedia of Database Systems, M. Tamer Özsu and Ling Liu, Eds., Springer, 2009, ISBN 978-0-387-35544-3 (print), ISBN 978-0-387-39940-9 (online), doi:10.1007/978-0-387-39940-9_629. Invited.

Vebjorn Ljosa and Ambuj K. Singh: "Probabilistic Querying and Mining of Biological Images," in *Managing and Mining Uncertain Data*, Charu C. Aggarwal, Ed., p. 461–482, Springer, February, 2009, ISBN 978-0-387-09689-6. Invited.

Vebjorn Ljosa and Anne E. Carpenter: "High-throughput screens for fluorescent dye discovery," *Trends in Biotechnology*, 26(10), p. 527–530, 2008, doi:10.1016/j.tibtech.2008.06.008. Peer-reviewed.

Ambuj K. Singh, Arnab Bhattacharya, and Vebjorn Ljosa: "Current challenges in bioimage database design," in *Proceedings of the International Workshop on Bioimage Data Mining and Informatics, IEEE Computational Systems Bioinformatics Conference (CSB)*, p. 375–379, August, 2005, doi:10.1109/CSBW.2005.47. Invited.

Conference presentations

Vebjorn Ljosa, Piyush B. Gupta, Thouis R. Jones, Eric S. Lander, and Anne E. Carpenter: "Large-scale learning of cellular phenotypes from images," in RECOMB Systems Biology, 2009.

Chao-I Chen, Xiaofang Lei, Vebjorn Ljosa, Thomas A. McCloskey, Chiou-Wei Tsai, S. Rao Jammalamadaka, Joel H. Rothman, and Ambuj K. Singh: "A Bayesian approach toward analyzing the network of genes involved in *C. elegans* programmed cell death," West Coast Worm Meeting, 2004.

Invited lectures

"Automatic quantification of subtle cellular phenotypes in microscopy-based high-throughput experiments," Fourth International Workshop on Machine Learning in Systems Biology (MLSB), Edinburgh, Scotland, 2010-10-15/--16. To appear.

"High-content screens for complex and subtle phenotypes," Bio-IT World Conference & Expo 2010 Workshop on Imaging Informatics, 2010-04-20.

"Quantifying complex and subtle phenotypes for image-based screening," Target Discovery World Congress South San Francisco, 2009-08-05.

"Scoring diverse cellular morphologies in image-based screens with iterative feedback and machine learning," Millennium – The Takeda Oncology Co., Cambridge, MA, 2009-06-01.

"Scoring diverse cellular morphologies in image-based screens with iterative feedback and machine learning," Society for Biomolecular Sciences 15th Annual Conference & Exhibition, Lille, France, 2009-04-29.

"Managing probabilistic data: Toward data-driven research in microscopy-based fields," Broad Institute of MIT and Harvard, 2007-04-05.

"Toward data-driven research on biological images: challenges for image analysis, data mining, machine learning, and databases ," University of Iowa, 2006-08-25.