Research Statement

Jeffrey Mark Siskind
http://engineering.purdue.edu/~qobi
http://upplysingaoflun.ecn.purdue.edu/~qobi/cccp/

Tuesday 13 December 2016

A unique aspect of Prof. Siskind's research leadership is that his expertise across multiple disciplines has allowed him to translate significant knowledge gleaned from research on human perception, processing, and learning into methods applied to computational vision and neurobiology beyond traditional approaches. For example, his pioneering work on recognizing action in video by analyzing the state changes in force-dynamic relations, published between 1991 and 2003, was motivated by the work of linguistics and psychologists like Baillargeon, Finke, Freyd, Keeble, Kestenbaum, Leslie, Michotte, McCloskey, Shephard, Spelke, and Talmy. It was among the earliest computational work on action recognition and still is the only work that is consistent with what we know about human perception and is so advanced that it has not yet been widely adopted by the computer-vision community. His other pioneering work on video action recognition, modeling action as time-series of changes in spatial relations between event participants, was motivated by speech processing. He was the first to apply continuous hidden Markov models, a technology well established in the field of speech recognition at the time, to video action recognition. This work on video action recognition, published between 1991 and 2002, presaged and predated the explosion of interest in action recognition within the computer-vision community by ten to fifteen years. As another example, his pioneering work on computational models of child language acquisition, published between 1990 and 1996, synthesized precise algorithmic methods from machine learning with knowledge of how children learn language, theorized by linguists like Grimshaw and Pinker, and obtained from psychology experiments conducted by psychologists like Gleitman, also presaged and predated the subsequent explosion of research interest in learning-based methods in computational linguistics by a decade or more. Further, his work on perceptual organization, published between 2001 and 2005, synthesized algorithmic methods from graph theory with Gestalt perceptual psychology, yielding the first computationally efficient and scale invariant algorithms for segmentation and contour completion. Moreover, he pioneered the whole enterprise of computationally grounding language in perception—today embodied in the fields of image and video captioning, and the emerging field of visual Q&A—starting with his thesis work in 1991 continuing through his recent best-paper awards in 2013 and 2015, his patent in 2015, and manuscripts currently in review. This work presaged and predated interest in image and video captioning, and visual Q&A, by a quarter of a century. It still is the only work within the field of computer vision that is consistent with what the fields of linguistics, psychology, and neuroscience know about human semantic processing, as it models the processes by which phrasal and sentential meaning is formed by linking the meanings of predicates to the meanings of arguments. Finally, his current work is the only that applies fMRI to human subjects performing the same tasks of action recognition and video captioning as being investigated by the computer-vision and natural-language processing communities, comparing the performance of current computer-vision and natural-language-processing methods with human brain processing.

From 2004 to 2010, Prof. Siskind worked on synthesizing the lambda calculus with the differential calculus. This work sits at the intersection of the fields of Automatic Differentiation (AD) and Function Programming (FP). In a seminal 1940 paper, Alonzo Church, the thesis advisor for Alan Turing, introduced the lambda calculus as a model of computation, laying out the foundation for all of computer science, and more specifically programming languages and compilers. The lambda calculus codified the notion of operators, prevalent in mathematics since the development of calculus, as higher-order functions. Church motivated this work by illustrating the derivative as a quintessential higher-order function. Over the subsequent seventy years, the techniques of automatic differentiation were developed within the field of scientific computation without ever introducing the derivative as an operator and higher-order programming languages were developed within the field of programming languages and compilers without ever introducing a higher-order function for derivatives. Until Prof. Siskind's work, together with Barak Pearlmutter, no one formulated a formal theory of computation that included a derivative operator. This theory is unique in exhibiting three properties: the derivative operator applies to the entire formal language, the syntactic range of the derivative transformation is a subset of its domain, allowing derivatives of derivatives, and the operators can properly nest, allowing one to take the derivative of a function that takes the derivative of another function. This allows nested minimax optimization to model competing and/or cooperative agents interacting with an environment, and as such is useful in most areas of scientific inquiry, ranging from economics to cognitive science.

Citation counts are from Google Scholar on Tuesday 1 November 2016. Impact factors are from JCR (2014). Only the 26 most cited papers are listed.

VLSI Design Automation Lead the development of the MacPitts silicon compiler, the first system that mapped behavioral descriptions of digital systems to custom VLSI layout. This presaged later systems, such as VHDL and Verilog, by two decades, that have become the cornerstone of the entire semiconductor industry. This work was licensed by MIT to GTE, and has become the foundation of products sold by GTE, Silc Technologies, Inc., Recal Redac Inc., Viewlogic Inc., and Mentor Graphics, Inc. The methods pioneered with MacPitts are now taught as part of the standard ECE curriculum throughout the world.

J.M. Siskind, J.R. Southard, and K.W. Crouch, 'Generating Custom High Performance VLSI Designs from Succinct Algorithmic Descriptions,' *Proceedings of the Conference on Advanced Research in VLSI*, pp. 28–40, January 1982, oral 21/82 (25%), 80 citations.
http://engineering.purdue.edu/~qobi/papers/arv1982.pdf

Child Language Acquisition Developed the first mathematical model of how children begin the process of acquiring language via cross-situational learning. Gathered the Brent-Siskind corpus, which was, at the time, the largest collection of maternal speech to infants in the Child Language Data Exchange System (CHILDES) with orthographic transcription time-aligned to the speech. Demonstrated that early child vocabulary is influenced more by isolated words than multiple-word utterances.

J.M. Siskind, 'Acquiring Core Meanings of Words, Represented as Jackendoff-Style Conceptual Structures, From Correlated Streams of Linguistic and Non-Linguistic Input,' *Proceedings of the Twenty Eighth Annual Meeting of the Association for Computational Linguistics (ACL)*, pp. 143–156, June 1990, oral, 36 citations.
http://engineering.purdue.edu/~qobi/papers/acl90.ps.Z

J.M. Siskind, 'Lexical Acquisition in the Presence of Noise and Homonymy,' *Proceedings of the Twelfth National Conference on Artificial Intelligence (AAAI)*, pp. 760–766, July 1994, oral 222/780 (28%), 28 citations.
http://engineering.purdue.edu/~qobi/papers/aaai94.ps.Z

J.M. Siskind, 'A Computational Study of Cross-Situational Techniques for Learning Word-to-Meaning Mappings,' *Cognition*, 61(1–2):39–91, October–November 1996, impact factor 3.479. Also appeared in *Computational Approaches to Language Acquisition*, M.R. Brent, ed., Elsevier, pp. 39–91, 1996, 477 citations.
http://engineering.purdue.edu/~qobi/papers/cognition1996.pdf

J.M. Siskind, 'Learning Word-to-Meaning Mappings,' in *Models of Language Acquisition: Inductive and Deductive Approaches*, P. Broeder and J. Murre ed., Oxford University Press, chapter 7, pp. 121–153, July 2000, 35 citations.
http://engineering.purdue.edu/~qobi/papers/cmla96.ps.Z

M.R. Brent and J.M. Siskind, 'The Role of Exposure to Isolated Words in Early Vocabulary Development,' *Cognition*, 81(2):B33–B44, September 2001, impact factor 3.479. Also available as Technical Report 99-107, NEC Research Institute, Inc., July 1999. Revised as Technical Report 2000-067R, NEC Research Institute, Inc., May 2000, 369 citations.
http://engineering.purdue.edu/~qobi/papers/cognition2001.pdf

Action Recognition Pioneered systems to recognize action in video a decade before it became mainstream. Developed the first system to recognize action in video using continuous Hidden Markov Models to model the changing spatial relations between participants during an action. This work was presented as an oral at the European Conference on Computer Vision (ECCV) 1996 and included in the video proceedings (as well as the conference proceedings), probably the first-ever video proceedings of a conference. Developed the first system to recognize action in video using temporal logic to model by the changing force-dynamic relations between participants during an action.

J.M. Siskind and Q. Morris, 'A Maximum-Likelihood Approach to Visual Event Classification,' *Proceedings of the Fourth European Conference on Computer Vision (ECCV)*, pp. 347–360, April 1996, oral 43/328 (13%), 84 citations.
http://engineering.purdue.edu/~qobi/papers/eccv96a.ps.Z

R. Mann, A.D. Jepson, and J.M. Siskind, 'The Computational Perception of Scene Dynamics,' *Proceedings of the Fourth European Conference on Computer Vision (ECCV)*, pp. 528–539, April 1996, poster 123/328 (37%), 44 citations.
http://engineering.purdue.edu/~qobi/papers/eccv96b.ps.Z

R. Mann, A.D. Jepson, and J.M. Siskind, 'The Computational Perception of Scene Dynamics,' *Computer Vision and Image Understanding (CVIU)*, 65(2):113–128, February 1997, impact factor 1.540, 79 citations.
http://engineering.purdue.edu/~qobi/papers/cviu1997.pdf

J.M. Siskind, 'Visual Event Classification via Force Dynamics,' *Proceedings of the Seventeenth National Conference on Artificial Intelligence (AAAI)*, pp. 149–155, August 2000, oral 143/432 (33%). Also available as Technical Report 2000-007, NEC Research Institute, Inc., January 2000. Revised as Technical Report 2000-047R, NEC Research Institute, Inc., April 2000, 54 citations.
http://engineering.purdue.edu/~qobi/papers/aaai2000.pdf

A.P. Fern, R.L. Givan, and J.M. Siskind, 'Specific-to-General Learning for Temporal Events with Application to Learning Event Definitions from Video,' *Journal of Artificial Intelligence Research (JAIR)*, 17:379–449, December 2002, impact factor 1.257, 56 citations.
http://engineering.purdue.edu/~qobi/papers/jair2002.pdf

J.M. Siskind, 'Reconstructing Force-Dynamic Models from Video Sequences,' *Artificial Intelligence (AIJ)*, 151(1–2):91–154, December 2003, impact factor 3.371, 39 citations.
http://engineering.purdue.edu/~qobi/papers/aij2003.pdf

Y. Cao, D. Barrett, A. Barbu, N. Siddharth, H. Yu, A. Michaux, Y. Lin, S. Dickinson, J.M. Siskind, and S. Wang, 'Recognize Human Activities from Partially Observed Videos,' *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2658–2665, Portland, OR, 25–27 June 2013, poster 472/1870 (25.2%), 51 citations.
http://engineering.purdue.edu/~qobi/papers/cvpr2013.pdf

Perceptual Organization Lead the development of Ratio Cut, the first polynomial time, globally optimal, scale invariant, graph based, image segmentation algorithm. Participated in the development of Ratio Contour, the first polynomial time, globally optimal, graph based, contour completion algorithm that encoded the Gestalt Laws of proximity, continuity, and closure. Led the development of Spatial Random-Tree Grammars and the Center-Surround algorithm, extending probabilistic context-free grammars and the inside/outside algorithm to multiple dimensions for learning grammars to parse images.

S. Wang and J.M. Siskind, 'Image Segmentation with Minimum Mean Cut,' *Proceedings of the Eighth International Conference on Computer Vision (ICCV)*, pp. 517–524, July 2001, poster 205/596 (34%). Also available as Technical Report 2000-169, NEC Research Institute, Inc., December 2000, 71 citations.
http://engineering.purdue.edu/~qobi/papers/iccv2001.pdf

S. Wang and J.M. Siskind, 'Image Segmentation with Ratio Cut,' *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 25(6):675–690, June 2003, impact factor 5.781, 323 citations.
http://engineering.purdue.edu/~qobi/papers/pami2003.pdf

S. Wang, T. Kubota, J.M. Siskind, and J. Wang, 'Salient Closed Boundary Extraction with Ratio Contour,' *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 27(4):546–561, April 2005, impact factor 5.781, 155 citations.
http://engineering.purdue.edu/~qobi/papers/pami2005.pdf

W. Wang, I. Pollak, T.-S. Wong, C.A. Bouman, M.P. Harper, and J.M. Siskind, 'Hierarchal Stochastic Image Grammars for Classification and Segmentation,' *IEEE Transactions on Image Processing (TIP)*, 15(10):3033–3052, October 2006, impact factor 3.625, 40 citations.
http://engineering.purdue.edu/~qobi/papers/tip2006.pdf

Programming Languages and Compilers Developed the Stalin compiler for Scheme, the first compiler to perform whole-program polyvariant flow analysis. Stalin is the highest performing Scheme compiler and one of the highest performing compilers for any language. Developed the Stalingrad compiler, the first compiler for a functional programming language that incorporates nestable higher-order automatic differentiation (AD). Stalingrad migrates run-time reflective source-to-source transformation to compile time using polyvariant flow analysis to generate extremely fast code for computing derivatives.

J.M. Siskind and D.A. McAllester, 'Nondeterministic Lisp as a Substrate for Constraint Logic Programming,' *Proceedings of the Eleventh National Conference on Artificial Intelligence (AAAI)*, pp. 133–138, July 1993, oral 135/>500 (<27%), 91 citations.
http://engineering.purdue.edu/~qobi/papers/aaai93.ps.Z

J.M. Siskind and D.A. McAllester, 'Screamer: A Portable Efficient Implementation of Nondeterministic Common Lisp,' Technical Report IRCS-93-03, Institute for Research in Cognitive Science, University of Pennsylvania, 1993, 44 citations.
http://engineering.purdue.edu/~qobi/papers/ircs-93-03.ps.Z

J.M. Siskind, 'Flow-Directed Lightweight Closure Conversion,' Technical Report 99-105, NEC Research Institute, Inc., July 1999. Revised as Technical Report 99-190R, NEC Research Institute, Inc., December 1999, 30 citations.
http://engineering.purdue.edu/~qobi/papers/fdlcc.pdf

J.M. Siskind and B.A. Pearlmutter, 'Nesting Forward-Mode AD in a Functional Framework,' *Higher-Order and Symbolic Computation (HOSC)*, 21(4):361–376 December 2008, 29 citations.
http://engineering.purdue.edu/~qobi/papers/hosc2008.pdf

Grounding Language in Vision and Robotics Pioneered the integration of computer vision, natural language processing, and robotics. Awarded the best-paper award at the Annual Meeting of the Association for Computational Linguistics (ACL) 2013 for this work. A paper on this work is included in the award-winning paper track of the Journal of Artificial Intelligence Research (JAIR). The methods developed over three and a half decades allow multi-directional and multimodal learning and inference across visual perception, robotic action, and linguistic description, including production of sentences that describe video, video search using sentential queries, production of sentences that describe mobile robot navigation, commanding mobile robot navigation using natural language, driving robotic game play from natural-language instructions, and robotic learning of game play through visual observation.

J.M. Siskind, 'Naive Physics, Event Perception, Lexical Semantics and Language Acquisition,' Ph.D. thesis, Artificial Intelligence Laboratory, MIT, January 1992, 93 citations.
http://engineering.purdue.edu/~qobi/papers/phd.ps.Z

J.M. Siskind, 'Grounding Language in Perception,' *Artificial Intelligence Review*, 8(5–6):371–391, 1995, impact factor 2.111, 101 citations.
http://engineering.purdue.edu/~qobi/papers/aireview1995.pdf

J.M. Siskind, 'Grounding the Lexical Semantics of Verbs in Visual Perception Using Force Dynamics and Event Logic,' *Journal of Artificial Intelligence Research (JAIR)*, 15:31–90, August 2001, impact factor 1.257. Also available as Technical Report 2000-105, NEC Research Institute, Inc., July 2000, 253 citations.
http://engineering.purdue.edu/~qobi/papers/jair2001.pdf

A. Barbu, A. Bridge, Z. Burchill, D. Coroian, S. Dickinson, S. Fidler, A. Michaux, S. Mussman, N. Siddharth, D. Salvi, L. Schmidt, J. Shangguan, J.M. Siskind, J. Waggoner, S. Wang, J. Wei, Y. Yin, and Z. Zhang, 'Video in sentences out,' *Proceedings of the 28th Conference on Uncertainty in Artificial Intelligence (UAI)*, pp. 102–112, Catalina, CA, 15–17 August 2012, oral 24/304 (8%), 68 citations.
http://engineering.purdue.edu/~qobi/papers/uai2012.pdf

H. Yu and J.M. Siskind, 'Grounded Language Learning from Video Described with Sentences,' *Proceedings of the Fifty First Annual Meeting of the Association for Computational Linguistics (ACL)*, pp. 56–63, Sofia, Bulgaria, 4–9 August 2013, oral 175/662 (26.4%), 81 citations, **best paper award**.
http://engineering.purdue.edu/~qobi/papers/acl2013.pdf

Neuroscience Pioneered the study of human language grounding in visual perception through fMRI. Presented paper at the senior-member track of the Conference on Artificial Intelligence (AAAI) 2015 on this work. Developed methods to study and quantify the compositional nature of human semantic processing in both vision and language, to quantify the degree to which such semantic processing is shared across both vision and language, and to quantify the degree to which such semantic processing is share across different subjects. Can show subjects sentence pairs and read out those sentences from brain-scan data. Can show subjects video clips depicting two simultaneous human actions and read out sentences that describe such visual percepts from brain-scan data. This work can correctly determine *who did what to what* when there are multiple *who*s, multiple *did what*s, and multiple *to what*s taking place simultaneously, associating the correct *who*, *did what*, and *to what*, for each pair of activities.