CS Community Drive launched; Why it matters

We know that you, our CS alumni, are strong and capable individuals—nearly 3,000 all over the world who are innovating, creating, and leading in a wide range of career pursuits. Although different in age and location, you share a common bond that helped to shape who you are today. And together with the dedicated CS faculty and staff, we are a Community. Now that we're 50 years old, we think it is time we took on the task of making the CS Community more connected, stronger, and more relevant for our alumni.

There are some easy ways for you to help us build our CS Community into the best it can be. Please consider one or all of the following ways you can help:

1. **Share your email address.** As of this writing, we are missing 900 alumni e-mail addresses. Please submit or update your email address so that we can keep you up-to-date with CS news and events. **www.cs.umass.edu/forms/email-address-form**.

2. **Attend a CS alumni event.** Reminisce with old friends, catch up with faculty and make new connections. Upcoming events include our Sixth Annual Outstanding Achievement and Advocacy Awards Banquet on May 2, 2014 (**www.cs.umass.edu/oaa2014**) and alumni socials planned for the east and west (CA Bay area) coasts.

3. **Give us news and updates** about your career and life to be included in *Significant Bits*. You can send them to Jean Joyce at jean@cs.umass.edu.

And finally,

4. **Donate in any amount to CS.** We know it's annoying to be asked for money. We know you probably get asked A LOT. But it really is important that you consider giving to CS. In order for us to remain competitive, we should be offering more scholarships to CS students. And, one of the factors used in determining our School's annual ranking is the Alumni Participation Rate (APR). The more alumni who donate, the higher the APR. The value of your CS degree increases when UMass and CS achieve high rankings.

Distinguished Professor Jim Kurose, chair of the CS 50th celebration committee, said it best: "Every contribution really *is* valued. When faculty and staff look through our list of donors in each issue of *Significant Bits* (and we *do*), it's really great to see the names of students we taught or knew. It tells us that they really value and appreciate their time with us. It means a lot to us and to UMass Amherst."

Steve Vinter (left) and Jim Kurose

Matching gift challenge

Thanks to CS alum Steve Vinter (Ph.D. '85), your gift to Computer Science during our 50th anniversary year will be doubled! He has agreed to match gifts made before July 1, 2014 (up to $50,000) in order to encourage fellow alumni to donate. "Now is the time YOU can help us make our wonderful school an even greater place," says Vinter. "Get involved! Let's join together and build the best foundation for our community that we can."

Why I give

No matter how large or how small, every single gift to CS is appreciated and put to good use. Here is why one CS alum chose to give…

In the mid-eighties, while taking a graduate course under Prof. Krithivasan Ramamritham, I read an IBM 1979 technical report, "Notes on Database Operating Systems" – precursor to the bible on Transaction Processing. Since then I've been hooked on building database systems. Later, at DEC and Microsoft, I was fortunate to get to know Jim Gray. As many of you know, he was lost at sea in 2007.

I set up the Jim Gray Scholarship at UMass Amherst to support first-year CS graduate students pursuing systems research. I did it because I want to encourage the next generation to join our field and keep it thriving. But I also did it because I want to honor and remember the sage of our science.

Hanuma Kodavalla (MS '88)

Lifetime email addresses for CS alums

We are pleased to announce that Computer Science alum email forwarding addresses are now available for any of our UMass Amherst CS alums. Although you might change employers or Internet providers, your CS alum email address will always stay the same and will forward your email to whatever address you choose. The email address will generally be firstname.lastname@alum.cs.umass.edu. To sign up for your email forwarding address, go to **www.cs.umass.edu/lifetime-email-forwarding**.

Beveridge and Draper team up on vision research

Bruce Draper (left) and Ross Beveridge

Colorado State University (CSU) Computer Science Professors Ross Beveridge and Bruce Draper did their Ph.D. work in the UMass Amherst Computer Vision Group, under the joint direction of Professors Ed Riseman and Al Hanson, and both graduated with Ph.D.s in 1993. The pair currently codirect the Computer Vision Research Group at CSU, following in the footsteps of Riseman and Hanson, leading the next generation of vision researchers.

At UMass Amherst, Beveridge and Draper saw firsthand an unusual model for academic research: namely, two good friends with very different personalities jointly coordinating a research group with multiple different projects. "At UMass, graduate students in the Vision group quickly learned that they didn't have one advisor, they had two," says Beveridge. "Whether Ed or Al was leading a particular project, the other was always following the work carefully and would quickly jump in with ideas of his own. Graduate students working in the CSU Computer Vision Group today learn a similar lesson, and the group is stronger for it."

Major accomplishments by Beveridge's and Draper's CSU Computer Vision Group include advanced analysis of algorithms that perform human face recognition and real-time activity recognition in video. In face recognition, the group has taken a leading role in the evaluation of algorithms, working closely since 2003 with Dr. Jonathon Phillips at NIST on a series of public face recognition challenge problems. In joint work with Geof Givens in the CSU Statistics Department, they have carried out the most detailed studies of interacting factors that influence face recognition algorithm performance.

Beveridge's and Draper's team also releases and maintains a suite of open source software packages that support labs around the world that want common face recognition baseline algorithm implementations and software for carrying performance studies according to standardized protocols. Most recently the researchers have taken the lead in a competition associated with the 2014 International Joint Conference on Biometrics where different labs and groups will work on a common video face recognition dataset.

Beveridge and Draper also work closely with CSU Mathematics Professors Michael Kirby and Chris Peterson on problems at the intersection of computer vision and algebraic topology. This work blends mathematics with practical utility in some surprising ways, including fundamental advancements in how to approach the problem of recognizing people's actions in video.

Their Vision Group is also known for its work on biomimetic approaches to computer vision, with an emphasis on unsupervised learning. "At its heart, this approach represents a commitment to developing systems that learn important internal representation without detailed guidance by people," says Draper. To express this in practical terms, the common approach to many computer vision tasks today is to require humans to hand label thousands of images, e.g. "That is a cat. That is a truck." This is tedious, and ultimately visually intelligent systems should not require such extraordinary hand-holding.

Recently, Beveridge's and Draper's group's work on biomimetic approaches to vision has been combined with the work on high-dimensional manifolds in the context of real-time activity recognition in video. This work was supported by the DARPA Minds Eye Program. A major component of their work was incorporated into a larger system that was then demonstrated by iRobot. The system learned quickly and then recognized actions such as running, walking, picking objects, carrying objects, etc. A video highlight of this effort is available at: **www.cs.colostate.edu/~draper/.**

Beveridge, Draper, and their team are now merging the video understanding and face recognition work. More generally, they remain committed to the same goals set out by Riseman and Hanson at UMass Amherst: machines should be using their eyes to watch, should be making sense of what they see, and then should be stepping in to help people in the myriad tasks that arise in all aspects of life.

A screen capture from a video where software developed by CSU and iRobot recognizes that a person is carrying something - in this case a shovel

We're on Facebook and LinkedIn

Keep up-to-date on the School's latest events and announcements. Join us on

Facebook (UMASS CS)
and
LinkedIn (UMass Amherst Computer Science)

John E. and Alice M. Flynn Scholarship established

Thanks to generous contributions from CS alum Daniel E. Flynn (BS '87), the **John E. and Alice M. Flynn Scholarship** was established in 2013 to honor his parents. The Scholarship will provide support to undergraduate students majoring in Computer Science. Preference will be given to students with financial need. According to Daniel, both John and Alice were "blue collar, hard-working, good people" with six children and eight grandchildren. They were lifelong residents of Massachusetts. Alice passed away in 1997 and John in 2007.

For more information or to contribute to the School of Computer Science, contact Jennifer Cooper at 413-545-2771 or jcooper@cns.umass.edu. You can also make a gift online at **www.cs.umass.edu/donate**, or by check made out to UMass Amherst Computer Science and mailed to Donna Falcetti, School of Computer Science, UMass Amherst, 140 Governors Drive, Amherst, MA 01003-9264.

Join us on May 2, 2014

The sixth annual Outstanding Achievement and Advocacy (OAA) Awards banquet will be held on Friday, May 2, 2014 in the Massachusetts Room of the UMass Amherst Mullins Center. During the banquet, awards will be presented to recognize the remarkable accomplishments of graduates of the School's degree programs and to acknowledge the support of important friends of the school. Current student awards will also be presented. For details and to register for the event, go to: **www.cs.umass.edu/oaa2014.**

CIIR west coast alum reunion

In October, Distinguished Professor Bruce Croft gathered for dinner with Center for Intelligent Information Retrieval (CIIR) alums and current students/staff during the CIKM conference in San Francisco, CA.

Front Row (left to right): Shiri Dori-Hacohen, Laura Dietz, Weize Kong, Xing Yi (Ph.D. '11), Wei Li (Ph.D. '07);
2nd Row: Jay Ponte (Ph.D. '98), Vanessa Murdock (Ph.D. '06), Jiwoon Jeon (Ph.D. '07), Yun Zhou (Ph.D. '08), Fernando Diaz (Ph.D. '07), Jie Bing (M.S. '13), Jangwon Seo (Ph.D. '11);
3rd Row: Michael Bendersky (Ph.D. '12), Jeff Dalton, Xiaobing Xue (Ph.D. '12), Mostafa Keikha, Pranav Mirajkar (M.S. '13), Trevor Strohman (Ph.D. '08), Greg Druck (Ph.D. '11);
Back Row: Jinyoung Kim (Ph.D. '12), Anton Leuski (Ph.D. '01), Paul Ogilvie (B.S. '00), Bruce Croft, and Ethem Can.

Brigadier General **Jody Daniels** (Ph.D. '97) will receive the 2014 campus Distinguished Alumni Service Award at the Massachusetts State House in April. She is a Deputy Director (J2 Intelligence and Knowledge Development) currently on active duty. More details on her award in our next issue.

Carla Brodley (Ph.D. '94), was named to the Defense Advanced Research Projects Agency (DARPA) Information Science and Technology study group in 2013 for a three-year term. The highly selective ISAT advisory group supports DARPA's technology offices, providing continuing and independent assessment of the state of advanced information science and technology, and their relationship to Department of Defense issues. Brodley is a Professor of Computer Science at Tufts University. She was Chair of the Department in the years 2010-2013.

Byron Wallace's (B.S. '06) thesis, "Machine Learning in Health Informatics: Making Better use of Domain Experts," was selected as the runner up for the 2013 ACM SIGKDD Doctoral Dissertation Award. He received his Ph.D. from Tufts University in 2012 and was advised by Carla Brodley. He is an Assistant Research Professor in the Department of Health Services and Public Policy, Center for Evidence-based Medicine, at Brown University.

CS alum and CSCF staffer **David Korpiewski** (B.S. '00) has started TransAerial, an aerial photography business (**www.transaerial.com**), which uses radio controlled drones to capture high definition video and pictures from the air. Korpiewski, a remote control model pilot, shot the photo of the CS building below.

Marc-Allen Cartright; *Query-Time Optimization Techniques for Structured Queries in Information Retrieval*; (James Allan, Advisor); Sept. 2013; Software Engineer, Google, Inc.

Information retrieval (IR) systems are evolving towards larger, more complicated queries. From an operational perspective, larger queries translate into an increasing computational cost to respond to a query. This causes an increasing tension in the trade-off between retrieval effectiveness and efficiency. This tension creates a strong need for optimization techniques to improve the efficiency of ranking with respect to these more complex retrieval models. This thesis presents three new optimization techniques designed to deal with different aspects of structured queries: 1) manipulation of interpolated subqueries, a commonly used query structure; 2) development of an alternative scoring formulation to make retrieval models more responsive to dynamic pruning techniques; and 3) introduction of delayed execution, which focuses on the class of queries that utilize term dependencies and term conjunction operations. In each case we empirically show that these optimizations can significantly improve query processing efficiency without impacting retrieval effectiveness.

Shane Clark; *The Security and Privacy Implications of Energy-Proportional Computing*; (Kevin Fu, Advisor); Sept. 2013; Research Scientist, BBN Technologies

The parallel trends of greater energy-efficiency and more aggressive power management are yielding computers that inch closer to energy-proportional computing with every generation. Energy-proportional computing, in which power consumption scales closely with workload, has unintended side effects for security and privacy. This thesis demonstrates the potential for system-level power analysis—the inference of a computer's internal states based on power observation at the "plug." It also examines which hardware components and software workloads have the greatest impact on information leakage. This work identifies the potential for privacy violations by demonstrating that a malicious party could identify which webpage from a given corpus a user is viewing. It also identifies constructive applications for power analysis, evaluating its use as an anomaly detection mechanism for embedded devices. Finally, this thesis includes modeling work that correlates AC and DC power consumption to pinpoint which components contribute most to information leakage.

Henry Feild; *Exploring Privacy and Personalization in Select Information Retrieval Applications*; (James Allan, Advisor); Sept. 2013; Assistant Professor, Dept. of Computer Science, Endicott College

The goal of this work is to explore the effects of personalization and privacy preservation methods on three information retrieval applications, namely search task identification, task-aware query recommendation, and searcher frustration detection. We pursue this goal by first introducing a novel framework called CrowdLogging for logging and aggregating data privately over a distributed set of users. We then describe several privacy mechanisms for sanitizing global data, including one novel mechanism based on differential privacy. We present a template for describing how local user data and global aggregate data are collected, processed, and used within an application, and apply this template to our three applications. We then introduce an open source system called CrowdLogger that implements the CrowdLogging framework and also serves as a platform for conducting in-situ user studies of search behavior, prototyping and evaluating information retrieval applications, and collecting labeled data.

Jacqueline Feild; *Improving Text Recognition in Images of Natural Scenes*; (Erik Learned-Miller, Advisor); Feb. 2014; Data Scientist, McGraw-Hill Education

This thesis develops methods for improving scene text recognition. We do this by incorporating new types of information into models and by exploring how to compose simple components into highly effective systems. First, we introduce two techniques for character recognition. We describe a character recognition system that incorporates similarity information in a novel way and a new language model that models syllables in a word to produce word labels that can be pronounced in English. Next we look at word recognition and we develop a new technique for segmenting text for these images called bilateral regression segmentation. We also introduce an open-vocabulary word recognition system that uses a very large web-based lexicon to achieve state of the art recognition performance. Lastly, we remove the assumption that words have been located and describe an end-to-end system that detects and recognizes text in any natural scene image.

Samuel Huston; *Indexing Proximity Based Dependencies for Information Retrieval*; (W. Bruce Croft, Advisor); Feb. 2014; Software Engineer, Google, Inc.

Research into term dependencies for information retrieval has demonstrated that dependency retrieval models are able to consistently improve retrieval effectiveness over bag-of-words models. However, the computation of term dependency statistics is a major efficiency bottleneck in the execution of these retrieval models. Further, despite the large number of published comparisons between dependency models and bag-of-words approaches, there has been a lack of direct comparisons between alternate dependency models. This thesis investigates the problem of improving the efficiency of dependency retrieval models without compromising the effectiveness benefits of the term dependency features. The major contributions presented in this thesis includes a systematic comparison of ad-hoc dependency models; the presentation and analysis of two new index data structures for term dependency data; and finally a comparison of the efficiency of these data structure for the most effective term dependency models identified by the systematic comparison.

Phillip Kirlin; *A Probabilistic Model of Hierarchical Music Analysis*; (David Jensen, Advisor); Feb. 2014; Assistant Professor, Dept. of Mathematics and Computer Science, Rhodes College

Schenkerian music theory supposes that Western tonal compositions can be viewed as hierarchies of musical objects. The process of Schenkerian analysis reveals this hierarchy by identifying connections between notes or chords of a composition that illustrate both the small- and large-scale construction of the music. We present a new probabilistic model of this variety of music analysis,

details of how the parameters of the model can be learned from a corpus, an algorithm for deriving the most probable analysis for a given piece of music, and both quantitative and human-based evaluations of the algorithm's performance. In addition, we describe the creation of the corpus, the first publicly available data set to contain both musical excerpts and corresponding computer-readable Schenkerian analyses. Combining this corpus with the probabilistic model gives us the first completely data-driven computational approach to hierarchical music analysis.

Chao Li; *Optimizing Linear Queries Under Differential Privacy*; (Gerome Miklau, Advisor); Sept. 2013; Software Engineer, Google, Inc.

Statistical data analysis on large collections of personal data can lead to fascinating results but also raises the privacy risk unwanted information disclosure. Differential privacy is a rigorous privacy definition that protects individuals' information during statistical data analysis. While it is straightforward to construct differentially private algorithms for many common tasks, methods to design error-optimal algorithms for most non-trivial tasks are still unknown. This thesis proposes the matrix mechanism, which answers sets of linear counting queries under differential privacy. Such queries cover the scope of many aggregation tasks, including count, sum and histogram. The thesis also contains an analysis of the matrix mechanism, including a closed-form error formulation and optimization programs to minimize the error. Further, it presents two algorithms. One based on the matrix mechanism and the other combining the matrix mechanism with other novel techniques, both of which answer various sets of queries with error lower than state-of-art algorithms.

Manjunath Narayana; *Probabilistic Models for Motion Segmentation in Image Sequences*; (Allen Hanson and Erik Learned-Miller, Advisors); Feb. 2014; Research Engineer, Metaio, Inc.

Motion segmentation, or labeling image pixels as moving or stationary, is an important task in computer vision. This thesis makes contributions towards segmentation with both stationary and moving cameras. For stationary cameras, we develop a probabilistic model that intuitively combines the various aspects of the problem in a system that is easy to interpret and to extend. For moving cameras, segmentation is commonly performed using the image plane motion of pixels, or optical flow. However, objects that are at different depths from the camera can exhibit different optical flows, causing a depth-dependent scene segmentation. We achieve a depth-independent segmentation that is consistent with real-world motion in the scene by using optical flow orientations instead of complete vectors. We propose a non-parametric probabilistic model that automatically estimates the number of motions and a rotation compensation algorithm that enables segmentation in a wide range of challenging hand-held camera videos

Scott Niekum; *Semantically Grounded Learning from Unstructured Demonstrations*; (Andrew Barto, Advisor); Sept. 2013; Postdoctoral Researcher, Carnegie Mellon University

Robots exhibit flexible behavior largely in proportion to their degree of semantic knowledge about the world. Such knowledge is often meticulously hand-coded for a narrow class of tasks, limiting the scope of possible robot competencies. For this reason, learning from demonstration (LfD) has become a popular alternative to traditional robot programming methods, aiming to provide a natural mechanism for quickly teaching robots. Unfortunately, LfD often yields little semantic knowledge about the world, and thus lacks robust generalization capabilities, especially for complex, multi-step tasks. To address this shortcoming of LfD, we present a series of algorithms that automatically detect and leverage repeated structure at multiple levels of abstraction in demonstration data, providing critical insights into task invariants, features of importance, and high-level task structure. This culminates in the discovery of semantically meaningful skills that are flexible and reusable, providing robust generalization and transfer in complex, multi-step robotic tasks.

James Partan; *Characterization and Network Consequences of Low Spreading Loss in Underwater Acoustic Networks*; (Brian Levine, Advisor); Sept. 2013; Research Engineer, Woods Hole Oceanographic Institution

The focus of this thesis is packet detection in interference in underwater acoustic wireless networks (UANs), and its role in the effectiveness of collision-avoidance medium-access control (MAC) protocols. Spreading-loss measures the decrease in received energy as a function of range, and determines the level of long-range interference. We present a new spreading model, the mixed-exponent spreading model, for UAN nodes using a matched-filter detector as a low-power wakeup detector. Under this model, there are distinct spreading-loss exponents for packet detection and interference. We validate this spreading model numerically, and with measurements of the spreading exponents from shallow-water experimental data. Our results suggest caution for use of the poorly grounded, but widely used, standard spreading model. We next analyze the effectiveness of collision-avoidance MAC protocols in UANs. The low spreading loss in UANs, in particular with the mixed-exponent spreading model, can lead to low collision-avoidance effectiveness compared with radio networks.

Ismet (Zeki) Yalniz; *Efficient Representation and Matching of Texts and Images in Scanned Book Collections*; (R. Manmatha, Advisor); Feb. 2014; Software Engineer, Amazon, A9.com

Several research problems are investigated over large collections of scanned books given their page images and corresponding optical character recognition (OCR) outputs. First, we propose general framework which can be used to efficiently align and compare the textual content of the books at various coarseness levels and even across languages. The framework uses the sequence of words which appear only once in the entire book ("the sequence of unique words") to represent the text. This approach is used for aligning long noisy texts, detecting partial duplicates and translations of books, and, aligning texts written in different languages. In the second part, the global font feature along with the letter sequence information is used for facilitating and/or improving text search in noisy page images using visual features. The effectiveness is demonstrated for books printed in different scripts for which there is no OCR engine available or the recognition accuracy is low.