



# VASC

---

## DEPARTMENTS & OFFICES

---

## NEWS & CALENDAR

---

## ACADEMICS & ADMISSIONS

---

## DIRECTORY, MAPS & VISITOR INFO

---

### Vision and Autonomous Systems Seminar

Training object class detectors typically requires a large set of images with objects annotated by bounding boxes, which are very time consuming to produce. In this talk I will present several schemes to reduce annotation time. These augment existing techniques for weakly supervised learning with a small amount of extra human annotation: (a) verifying bounding-boxes produced by the learning algorithm; (b) clicking on the object center; (c) response times measured during visual search. I will show that this extra annotation can go a long way: some of these schemes deliver detectors almost as good as those trained in a fully supervised setting, while reducing annotation time by about 10x. I will conclude with our effort to annotate part of the COCO dataset with a broad range of stuff classes. To this end we propose a specialized annotation protocol which leverages existing thing annotations to enable to efficiently label all stuff pixels.

—

[Vittorio Ferrari](#) is a Professor at the School of Informatics of the University of Edinburgh and a Research Scientist at Google, leading a research group on visual learning in each institution. He received his PhD from ETH Zurich in 2004 and was a post-doctoral researcher at INRIA Grenoble in 2006-2007 and at the University of Oxford in 2007-2008. Between 2008 and 2012 he was Assistant Professor at ETH Zurich, funded by a Swiss National Science Foundation Professorship grant. He received the prestigious ERC Starting Grant, and the best paper award from the European Conference in Computer Vision, both in 2012. He is the author of over 90 technical publications. He regularly serves as an Area Chair for the major computer vision conferences, he will be a Program Chair for ECCV 2018 and a General Chair for ECCV 2020. He is an Associate Editor of IEEE Pattern Analysis and Machine Intelligence. His current research interests are in learning visual models with minimal human supervision, object detection, and semantic segmentation.

**Host:** Olga Russakovsky

### Vision and Autonomous Systems Seminar - Cancelled

Consensus-based distributed learning seeks to find a general consensus of local learning models to achieve a global objective. Problems of this type arise in many settings, include distributed sensor networks, big data, as well as complex systems such as the human-cyber-physical systems, where either computational or physical constraints prevent traditional, centralized data analytics solutions. In this talk I will focus on the decentralized merits of distributed learning, taking it one step further from alternative parallel methods.

First, I will discuss a general distributed probabilistic learning framework based on alternating direction method of multipliers (ADMM) and show how it can be applied in computer vision algorithms which traditionally assume a centralized computational setting. We demonstrate that our probabilistic interpretation is useful in dealing with missing values, which is not explicitly handled in prior works. I will next present an extension of this approach to online decentralized probabilistic learning and will also show how the learning process can be accelerated by introducing new update strategies to the underlying optimization algorithm.

Finally, I will introduce our recent work on human crowd behavior estimation problem in the context of decentralized learning. I will demonstrate how the group trajectory estimation problem can be recast as a decentralized state estimation approach and can be, subsequently, augmented to include physics-data driven fusion. I will show that our approach can effectively reconstruct noisy, corrupted trajectories from off-the-shelf human trackers that could help human crowd analysis and simulation in the context of large cyber-physical systems.

---

[Vladimir Pavlovic](#) is a Professor at the Computer Science Department at Rutgers University. He received the PhD in electrical engineering from the University of Illinois in Urbana-Champaign in 1999. From 1999 until 2001 he was a member of research staff at the Cambridge Research Laboratory, Cambridge, MA. Before joining Rutgers in 2002, he held a research professor position in the Bioinformatics Program at Boston University. Vladimir's research interests include probabilistic system modeling, time-series analysis, statistical computer vision and data science. His work has been published in major computer vision, machine learning and pattern recognition journals and conferences. More information can be found on his [group](#).

*Sponsored in part by Disney Research.*

## Vision and Autonomous Systems Center Seminar

Physics-based modeling research in graphics has been consistently conscious of advances in modern parallel hardware, leveraging new performance capabilities to improve the scope and scale of simulation techniques. An exciting consequence of such developments is that a number of performance-hungry emerging applications, including computer-aided healthcare and medical training, can now hope to be accommodated in interactive systems. Nevertheless, while large-scale simulation for production-grade visual effects always had the option of clustering compute resources to keep up with growing needs, realtime or near-interactive applications face a more complex set of challenges. In fact, extracting competitive levels of efficiency out of modern parallel platforms is more often than not the result of cross-cutting interventions across the spectrum of theory, modeling, numerics and software engineering.

In this talk I will present a number of examples, mostly drawn from biomechanical modeling, virtual surgery and anatomical simulation tasks, where fresh perspectives on discretization, geometrical modeling, data-parallel programming or even the formulation of the governing PDEs for a physical system were instrumental in boosting parallel efficiency. Finally, I will discuss important lessons learned from simulations of human anatomy, and how those pertain to the design of solvers for computational physics at large, and particularly how they can boost the scale and efficiency of highly detailed fluid dynamics simulations.

---

[Eftychios Sifakis](#) is an Assistant Professor of Computer Sciences and (by courtesy) Mechanical Engineering and Mathematics at the University of Wisconsin-Madison. He obtained his Ph.D. degree in Computer Science (2007) from Stanford University. Between 2007-2010 he was a postdoctoral researcher in the University of California Los Angeles, with a joint appointment in Computer Science and Mathematics. His research focuses on scientific computing, physics based modeling and computer graphics. He is particularly interested in biomechanical modeling for applications such as character animation, medical simulations and virtual surgical environments. Eftychios has served as a research consultant with Intel Corporation, Walt Disney Animation Studios and SimQuest LLC, and is a recipient of the NSF CAREER award (2013-2018).

*Sponsored in part by Disney Research.*

## Vision and Autonomous Systems Center Seminar

Demonstrations from computer vision, such as the recent example of successful navigation without generating any 3D map (Zhu et al, 2016), are likely to have a profound influence on hypotheses about the type of representation that the brain uses to when faced with similar tasks. The goal of work in my lab is to find psychophysical evidence to help discriminate between rival models of 3D vision. The critical division is between models based on 3D coordinate frames and those that use something more like a graph of views.

I will present data from our virtual reality lab, where observers move freely and carry out simple tasks such as navigating to remembered locations or making judgments about the size, distance or direction of objects. We often manipulate the scene as participants move, e.g. expanding the world several-fold in all dimensions which participants fail to notice. In all cases, the data are difficult to explain under an assumption that the brain generates a single 3D reconstruction of the scene independent of the task. An alternative is that the brain stores something more like a graph of sensory states linked by actions (or, in fact, 'sensory+motivational' states, which is closely related to the embedding of sensory and goal information that Zhu et al adopt)

[Zhu, Mottaghi, Kolve, Lim, Gupta, Fei-Fei, Farhadi \(2016\)](#)

---

[Andrew Glennerster](#) studied medicine at Cambridge before doing his DPhil in Oxford in Experimental Psychology on human binocular stereopsis. He set up a virtual reality lab in the Physiology department in Oxford where he had Fellowships from the Medical Research

Council and the Royal Society. He continues to work on 3D vision in moving observers at the University of Reading where he is a Professor in the School of Psychology and Clinical Language Sciences.

*Sponsored in part by Disney Research.*

## Vision and Autonomous Systems Center Seminar

We study the problem of learning geometric and physical object properties from visual input. Inspired by findings in cognitive science that even infants are able to perceive a physical world full of dynamic content, we aim to build models to characterize physical and geometric object properties from synthetic and real-world scenes. In this talk, I will present some models we recently proposed for 3D shape recognition and synthesis, and for physical scene understanding. I will also present a newly collected video dataset of physical events.

---

**Jiajun Wu** is a third-year Ph.D. student at Massachusetts Institute of Technology, advised by Professor Bill Freeman and Professor Josh Tenenbaum. His research interests lie on the intersection of computer vision, machine learning, and computational cognitive science. Before coming to MIT, he received his B.Eng. from Tsinghua University, China, advised by Professor Zhuowen Tu. He has also spent time working at research labs of Microsoft, Facebook, and Baidu.

## Vision and Autonomous Systems Center Seminar

With 3D printing, geometry that was once too intricate to fabricate can now be produced easily and inexpensively. In fact, printing a object perforated with thousands of tiny holes can be actually cheaper and faster than producing the same object filled with solid material. The expanded designspace admitted by additive fabrication contains objects that can outperform traditional shapes and exhibit interesting properties, but navigating the space is a challenging task.

In this talk, I focus on two applications leveraging this design space. First, I discuss how to customize objects' deformation behaviors even when printing with a single material. By designing structures at the microscopic scale, we can achieve perfectly isotropic elastic behavior with a wide range of stiffnesses (over 10000 times softer than the printing material) and effective Poisson's ratios (both auxetic and nearly incompressible). Then, with an optimization at the macroscopic scale, we can decide where in the object to place these structures to achieve a user-specified deformation goal under prescribed forces.

Next I tackle a problem that emerges when using micro-scale structures: fragility, especially for the softer designs. I discuss how to efficiently analyze structures for their likelihood of failure (either brittle or ductile fracture) under general use. Finally, I show how to optimize a structure to maximize its robustness while still maintaining its elastic behavior.

---

**Julian Panetta** is a PhD candidate at NYU's Courant Institute, where he is advised by Denis Zorin. Julian is interested in simulation and optimal design problems, specifically focusing on applications for 3D printing. Before joining NYU, he received his BS in computer science from Caltech and did research at NASA's Jet Propulsion Lab.

*Sponsored in part by Disney Research*

## Vision and Autonomous Systems Center Seminar

Understanding and reasoning about our visual world is a core capability of artificial intelligence. It is a necessity for effective communication, and for question/answering tasks. In this talk, I discuss some recent explorations into visual reasoning to gain an understanding of how humans and machines tackle the problem. I'll also describe how algorithms initially developed for visual understanding can be applied to other domains, such as program induction.

---

**C. Lawrence Zitnick** is a research manager at Facebook AI Research, and an affiliate associate professor at the University of Washington. He is interested in a broad range of topics related to artificial intelligence including object recognition, the relation of language and imagery, and methods for gathering common sense knowledge. He developed the PhotoDNA technology used by Microsoft, Facebook, Google, and various law enforcement agencies to combat illegal imagery on the web. Before joining FAIR, he was a principal researcher at Microsoft Research. He received the PhD degree in robotics from Carnegie Mellon University.

## Vision and Autonomous Systems Center Seminar

Everyone has some experience of solving jigsaw puzzles. When facing ambiguities of assembling a pair of pieces, a common strategy we use is to look at clues from additional pieces and make decisions among all relevant pieces together. In this talk, I will show how to apply this common practice to develop data-driven algorithms that significantly outperform pair-wise algorithms. I will start with describing a computation framework for the joint inference of correspondences among shape/image collections. Then I will discuss how similar ideas can be utilized to learn visual correspondences.

---

**Qixing Huang** is an assistant professor at the University of Texas at Austin. He obtained his PhD in Computer Science from Stanford University and his MS and BS in Computer Science from Tsinghua University. He was a research assistant professor at Toyota Technological Institute at Chicago before joining UT Austin. He has also worked at Adobe Research and Google Research, where he developed some of the key technologies for Google Street View.

Dr. Huang's research spans the fields of computer vision, computer graphics, and machine learning. In particular, he is interested in designing new algorithms that process and analyze big geometric data (e.g., 3D shapes/scenes). He is also interested in statistical data analysis, compressive sensing, low-rank matrix recovery, and large-scale optimization, which provides theoretical foundation for his research. Qixing has published extensively at SIGGRAPH, CVPR and ICCV, and has received grants from NSF and various industry gifts. He also received the best paper award at the Symposium on Geometry Processing 2013.

## Vision and Autonomous Systems Center Seminar

Beginning with the philosophical and cognitive underpinnings of referring expression generation, and ending with theoretical, algorithmic and applied contributions in mainstream vision-to-language research, I will discuss some of my work through the years towards the ultimate goal of helping humans and computers to communicate. This will be a multi-modal, multi-disciplinary talk (with pictures!), aimed to be interesting no matter what your background is.

---

**Meg Mitchell** is currently a Senior Research Scientist in Google's Machine Intelligence Research in Seattle, WA. I work on advancing artificial intelligence in a way that is interpretable, understanding of art and literature, and respectful of user privacy. I work on vision-language and grounded language generation, focusing on how to help computers communicate based on what they can process. My work combines computer vision, natural language processing, social media, many statistical methods, and insights from cognitive science. I continue to balance my time between language generation, applications for clinical domains, and core AI research.

*Sponsored in part by Disney Research*

## Vision and Autonomous Systems Center Seminar

Structures and objects, captured in image data, are often idealized by the viewer. For example, buildings may seem to be perfectly straight, or repeating structures such as corn's kernels may seem almost identical. However, in reality, such flawless behavior hardly exists. The goal in this line of work is to detect the spatial imperfection, i.e., departure of objects from their idealized models, given only a single image as input, and to render a new image in which the deviations from the model are either reduced or magnified. Reducing the imperfections allows us to idealize/beautify images, and can be used as a graphic tool for creating more visually pleasing images. Alternatively, increasing the spatial irregularities allow us to reveal useful and surprising information that is hard to visually perceive by the naked eye (such as the sagging of a house's roof). I will consider this problem under two distinct definitions of idealized model: (i) ideal parametric geometries (e.g., line segments, circles), which can be automatically detected in the input image. (ii) perfect repetitions of structures, which relies on the redundancy of patches in a single image. Each of these models has led to a new algorithm with a wide range of applications in civil engineering, astronomy, design, and materials defects inspection.

---

**Tali Dekel** is currently a Research Scientist at Google, working on developing computer vision and computer graphics algorithms. Before Google, she was a Postdoctoral Associate at the Computer Science and Artificial Intelligence Lab (CSAIL) at MIT, working with Prof. William T. Freeman. Tali completed her Ph.D studies at the school of electrical engineering, Tel-Aviv University, under the supervision of Prof. Shai Avidan, and Prof. Yael Moses. Tali's Ph.D. focused on the use of multi-camera systems to solve classic and innovative tasks in computer vision and computer graphics including 3D structure and 3D motion estimation, content-geometry aware stereo retargeting, and photo sequencing (recovering temporal order of distributed image set). In her postdoc studies, she has been working on developing new algorithms that detect and visualize imperfections/irregularities in a single image. Her research interests include computer vision and graphics, geometry, 3D reconstruction, motion analysis, and image visualization.

## Upcoming Events [[More](#)]

THURSDAY, JULY 6, 2017 - 2:00PM

### [Robotics Thesis Defense](#)

MARYNEL VAZQUEL  
Reddy Conference Room 4405

Gates Hillman Centers

---

FRIDAY, JULY 7, 2017 - 11:00AM

### [Computer Science Thesis Oral](#)

DAVID THOMAS NAYLOR  
8102

Gates Hillman Centers

---

TUESDAY, JULY 11, 2017 - 3:30PM

### [Principals of Programming Seminar](#)

GABRIEL SCHERER  
Blelloch-Skees Conference Room 8115

Gates Hillman Centers

---

THURSDAY, JULY 13, 2017 - 12:00PM

### [Human-Computer Interaction Thesis Defense](#)

KELLY RIVERS  
ASA Conference Room 6115

Gates Hillman Centers

---

TUESDAY, JULY 18, 2017 - 9:00AM

### [Human-Computer Interaction Thesis Defense](#)

JENNIFER OLSEN  
Reddy Conference Room 4405

Gates Hillman Centers

---



## News in Brief [[More](#)]

03.28.17 --

Adarsh Prasad and Enxu Yan, Ph.D. students in [machine learning](#), and Akhil Chaturvedi, a graduate student in [biomedical engineering](#), won first place and a \$25,000 prize in the [Citadel/Citadel Securities Datathon](#), hosted by CMU's Department of Statistics March 25. The runner-up team of [math](#) major... [MORE](#) »

03.22.17 --