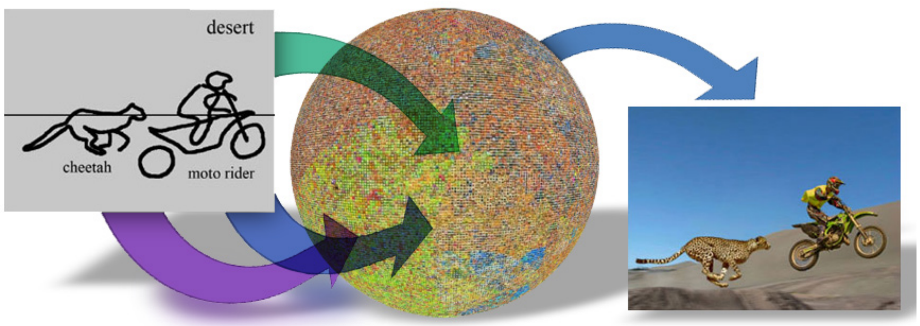
Data-Driven Computer Vision

Brown University Spring 2016

Instructor: Genevieve Patterson
Contact: gen@cs.brown.edu
Office Hours: CIT 551 MW 1-3pm



November 3, 2015

Course Description

This graduate seminar course investigates current research topics in data-driven object detection, scene recognition, and image-based graphics. We will examine data sources, features, and algorithms useful for understanding and manipulating visual data. We will pay special attention to methods that harness large-scale or Internet-derived data. There will be an overview of the current crowdsourcing techniques used to acquire massive image datasets. Vision topics such as scene understanding and object detection will be linked to graphics applications such as photo editing. These topics will be pursued through independent reading, class discussion and presentations, and projects involving current research problems in Computer Vision.

The goal of this course is to give students the background and skills necessary to perform research in image-based graphics and vision. Students should understand the strengths and weaknesses of current approaches to research problems and identify interesting open questions and future research directions. Students will hopefully improve their critical reading and communication skills, as well.

Course Requirements

Reading and Summaries

Students will be expected to read one paper for each class. For each assigned paper, students must write a two or three sentence summary and identify at least one question or topic of interest for class discussion. Interesting topics for discussion could relate to strengths and weaknesses of the paper, possible future directions, connections to other research, uncertainty about the conclusions of the experiments, etc. Reading summaries must be posted to the class blog by 11:59pm the day before each class. Students are encouraged to reply to other comments on the blog and help each other understand confusing aspects of the papers. The blog discussion will be the starting point for the class discussion. Students presenting a paper in class don't need to post a summary to the blog.

Class preparation and participation

All students are expected to take part in class discussions. Students are not required to fully understand a paper. We expect to work through the unclear aspects of a paper together in class.

Many of the papers covered in this course will have publicly available code and tutorials for running their systems and experiments. For these papers, students will be expected to run the basic versions of the paper's system. Students are not expected to re-implement the entire system or set of experiments. The purpose of these tutorial exercises is to familiarize students with running code written by other researchers. Students will be expected to identify strengths and weaknesses of the systems they attempt to run.

In class, students will be asked to explain strengths and weaknesses of the presented algorithms. At a minimum, students are expected to regularly volunteer analysis of performance, memory consumption, computational complexity, and algorithmic bias for the systems and algorithms discussed in class.

Attendance

Students are expected to attend all classes in order to participate in in-class discussion. Students are allowed 2 excused absences during the semester for any reason. Additional absences must have a medical or administrative excuse, and otherwise will decrease the in-class participation grade by 5%.

Presentation(s)

Depending on enrollment, students will lead the discussion of one or two papers during the semester. Ideally, students would implement some aspect of the presented material and perform experiments that help understand the algorithms. Presentations and all supplemental material should be ready one week before the presentation date so that students can meet with the instructor, go over the presentation, and possibly iterate before the in-class discussion. For the presentations it is fine to use slides and code from outside sources (for example, the paper authors), but students must be sure to give credit.

Semester projects

Students, with the help of course staff, will come up with research projects of current interest, which they will then complete by the end of the semester. Students will propose a research topic part way through the semester. After a project topic is finalized, students will meet occasionally with the instructor to discuss progress. Students will present their progress on their semester project twice during the course and the course will end with final project presentations. Students will also produce a conference-formatted write-up of their project. Project reports will be due one week before final presentations. Projects will be published on the the course web page. The ideal project is something with a clear enough direction to be completed in a couple of months, and enough novelty such that it could be published in a peer-reviewed venue with some refinement and extension.

Late Work

Neither reading summaries nor in-class paper presentations may be submitted late due to the timing of the course. If a student believes they will not be able to present a paper in class on the assigned date, they should reschedule their presentation with the instructor for a 10% deduction in grade. The semester project is due by the final day of class, and also cannot be submitted late.

Textbook

We will not rely on a textbook, although the free, online textbook ”Computer Vision: Algorithms and Applications” by Richard Szeliski is a helpful resource.

Grading

The final grade will be made up from

- 20% Reading summaries posted to class blog
- 20% Class preparation and participation
- – 10% Preliminary algorithm implementation/ Tutorial implementation
- – 10% In class discussion of each paper’s system’s performance
- 20% Research presentation(s)
- 40% Semester project

Schedule

Week #	Paper
	Introduction
1	The state of vision and graphics
	Image Representations
1	**80 million tiny images: a large dataset for non-parametric object and scene recognition,** A. Torralba, R. Fergus, W. T. Freeman. IEEE Transactions on Pattern Analysis and Machine Intelligence, vol.30(11), 2008.
1	**Object recognition from local scale-invariant features,** David Lowe, ICCV 1999.
2	**Video Google: A Text Retrieval Approach to Object Matching in Videos,** Sivic, J. and Zisserman, A. Proceedings of the International Conference on Computer Vision (2003).
2	**Beyond Bags of Features: Spatial Pyramid Matching for Recognizing Natural Scene Categories,** . Lazebnik, C. Schmid, and J. Ponce, CVPR 2006.
2	**Scene Completion Using Millions of Photographs,** James Hays, Alexei A. Efros. ACM Transactions on Graphics (SIGGRAPH 2007). August 2007, vol. 26, No. 3.
	Learned Representations, Deep Learning
3	**Caffe Deep Learning Tutorial at CVPR 2015,** Evan Shelhamer, Jeff Donahue, Jon Long, Yangqing Jia, and Ross Girshick. Associated paper: **Caffe: Convolutional architecture for fast feature embedding,** Y. Jia, E. Shelhamer, J. Donahue, S. Karayev, J. Long, R. Girshick, S. Guadarrama, and T. Darrell. In Proceedings of the ACM International Conference on Multimedia, pages 675 678. ACM, 2014.
3	**ImageNet Classification with Deep Convolutional Neural Networks.,** Alex Krizhevsky, Ilya Sutskever, Geoffrey E Hinton. NIPS 2012.
3	**Visualizing and Understanding Convolutional Networks,** Matthew D Zeiler, Rob Fergus. ECCV 2014.
4	**Rich feature hierarchies for accurate object detection and semantic segmentation,** Girshick, Ross, et al. Computer Vision and Pattern Recognition (CVPR), 2014.
	Crowdsourcing and Human Computation
4	**LabelMe: a Database and Web-based Tool for Image Annotation,** B. C. Russell, A. Torralba, K. P. Murphy, W. T. Freeman. International Journal of Computer Vision, 2008.
4	**ImageNet: A Large-Scale Hierarchical Image Database,** J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li and L. Fei-Fei. IEEE Computer Vision and Pattern Recognition (CVPR), 2009.
5	Project Status Updates.

5	Project Status Updates.
5	**Visual Recognition with Humans in the Loop,** Steve Branson, Catherine Wah, Florian Schroff, Boris Banenko, Peter Welinder, Pietro Perona, and Serge Belongie. ECCV 2010.
6	**The Multidimensional Wisdom of the Crowds,** Peter Welinder, Steve Branson, Pietro Perona, Serge Belongie. NIPS 2010.
6	**How do humans sketch objects?** Mathias Eitz, James Hays, and Marc Alexa. Siggraph 2012.
6	**Micro Perceptual Human Computation for Visual Tasks,** Yotam Gingold, Ariel Shamir, Daniel Cohen-Or. ACM Transactions on Graphics (ToG) 2012.
7	**Cascade: Crowdsourcing Taxonomy Creation,** Lydia Chilton, Greg Little, Darren Edge, Daniel Weld, and James Landay. Proceeding of the 2013 ACM Annual Conference on Human Factors in Computing Systems, pages 199-2008. ACM, 2013.
	Quizz: Targeted crowdsourcing with a billion (potential) users. Ipeirotis, Panagiotis G., and Evgeniy Gabrilovich. Proceedings of the 23rd international conference on World wide web. ACM, 2014.
7	**Microsoft COCO: Common Objects in Context,** Tsung-Yi Lin, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollr, and C. Lawrence Zitnick. ECCV 2014.
Leveraging Image Databases	
7	**Photo Clip Art,** Jean-Franois Lalonde, Derek Hoeim, Alexei A. Efros, Carsten Rother, John Winn and Antonio Criminisi. ACM Transactions on Graphics (SIGGRAPH 2007).
8	**Sketch2Photo: Internet Image Montage,** ACM SIGGRAPH ASIA 2009, ACM Transactions on Graphics. Tao Chen, Ming-Ming Cheng, Ping Tan, Ariel Shamir, Shi-Min Hu.
8	**AverageExplorer: Interactive Exploration and Alignment of Visual Data Collections,** Jun-Yan Zhu, Yong Jae Lee, Alexei Efros. Siggraph 2014.
Attribute-based Representations	
9	**Describing Objects by Their Attributes,** A. Farhadi, I. Endres, D. Hoiem, and D.A. Forsyth. CVPR 2009.
9	**Automatic attribute discovery and characterization from noisy web data,** Berg, Tamara L., Alexander C. Berg, and Jonathan Shih. Computer VisionECCV 2010. Springer Berlin Heidelberg, 2010. 663-676.
9	**The SUN Attribute Database: Beyond Categories for Deeper Scene Understanding,** Genevieve Patterson, Chen Xu, Hang Su, James Hays. IJCV 2014.
10	Project Status Updates.
10	Project Status Updates.
Learning Image Similarity and Context	

10	**Visual Turing test for computer vision systems,** Geman, Donald, et al. Proceedings of the National Academy of Sciences 112.12 (2015): 3618-3623.
11	**Visual concept learning: Combining machine vision and bayesian generalization on concept hierarchies,** Jia, Yangqing, et al. Advances in Neural Information Processing Systems (NIPS). 2013.
11	**Using very deep autoencoders for content-based image retrieval,** Krizhevsky, Alex, and Geoffrey E. Hinton. ESANN. 2011.
11	**Learning Deep Representations for Ground-to-Aerial Geolocalization,** Tsung-Yi Lin, Yin Cui, Serge Belongie, James Hays. CVPR 2015.
Structure from Motion	
12	**Photo tourism: Exploring photo collections in 3D,** Noah Snavely, Steven M. Seitz, Richard Szeliski. Siggraph 2006.
12	**First Person Hyperlapse Videos,** Johannes Kopf, Michael Cohen, Richard Szeliski. Siggraph 2014.
12	**Eulerian video magnification for revealing subtle changes in the world,** Wu, Hao-Yu, et al. ACM Trans. Graph. 31.4 (2012): 65.
Misc	
13	**A High Performance CRF Model for Clothes Parsing,** E Simo-Serra, S Fidler, F Moreno-Noguer, R Urtasun Computer Vision–ACCV 2014.
13	**Visual Concept Learning: Combining Machine Vision and Bayesian Generalization on Concept Hierarchies** Y Jia, J Abbott, J Austerweil, T Griffiths, T Darrell. NIPS 2013.
13	**Transient Attributes for High-Level Understanding and Editing of Outdoor Scenes,** Pierre-Yves Laffont, Zhile Ren, Xiaofeng Tao, Chao Qian, James Hays. Siggraph 2014.
14	Final Project Presentations.
14	Final Project Presentations

Acknowledgements

This course is closely modeled after the course "Data-Driven Vision and Graphics," taught at Brown by James Hays. Ideas for the organization and content of this course also came from many other researchers such as Svetlana Lazebnik, Kristin Grauman, Antonio Torralba, Derek Hoeim, and Alexei Efros.