Spring 2022, CS294-167 – Geometry and Learning for 3D Vision

Overview

This course gives a systematic introduction to the geometric principles and computational methods of recovering three-dimensional (3D) scene structure and camera motion from multiple, or a sequence of, two-dimensional (2D) images. The first part of the course provides a complete and unified characterization of *all fundamental geometric relationships* among multiple 2D images of 3D points, lines, planes, and symmetric structures etc., as well as the associated geometric reconstruction algorithms. Complementary to the geometry, the second part of the course introduces latest development in supervised or unsupervised *learning-based methods* for detecting and recognizing local features or global geometric structures (e.g. wireframes, planes, regular textures, symmetric objects) in 2D images, for robust and accurate 3D reconstruction. Although principles and methods introduced are fundamental and general, this course emphasizes applications in augmented reality and autonomous 3D mapping and navigation.

This course can be viewed as an advanced course in computer vision with a focus on 3D Vision. It can be taken as a follow-up of the Computer Vision course CS 280 or the 3D Image Processing course EE290T (offered by Zakhor). It is also designed for students who have taken an introductory Robotics course, say EECS 106, and wish to systematically learn about machine vision for purposes such as localization, mapping, and navigation. However, this course is entirely self-contained, necessary background knowledge in linear algebra, rigid-body motions, image formation, and camera models will be covered in the very beginning.

Administrative

Time and place: Wed 2:00PM - 4:00PM, 306 Soda

Instructor: Professor Yi Ma
Instructor email: `yima@eecs.berkeley.edu`.

Course webpage: We will use a piazza website to post lecture materials, homeworks, code examples, etc.

Computing tools: We will use the Matlab, Jupyter notebooks, and Google Colab environments for many of the in-lecture demos, examples and homeworks.

Prerequisites

Linear Algebra and Probability. Background in signal processing, optimization, and machine learning may allow you to appreciate better certain aspects of the course material, but not necessary all at once. If you're curious about whether you would benefit from this course, contact the instructor for details. The course is open to senior undergraduates, with consent from the instructor.

Textbooks and Resources

We will use the following textbook that covers the fundamental geometrical principles for 3D reconstruction from 2D images. (Students will be provided with copies of this textbook.)

- **Textbook:** "*An Invitation to 3-D Vision: From Images to Geometric Models*," Yi Ma, Stefano Soatto, Jana Kosecka, and Shankar Sastry, Interdisciplinary Applied Mathematics #26, Springer, 2004

We will also provide a few other related textbooks and references on the course website. For more recent developments on learning-based recognition of geometric features or structures for reconstruction purposes, we will utilize resources (papers, data, code etc.) from the public site:

- **Github:** "*Holistic 3D Reconstruction*," https://github.com/holistic-3d/holistic-3d-resources.

Grades

The course will be graded based on class participation (10%), homework (20%) and a course project (70%). For the course project, you can work on a topic of your choice – experimental, theoretical, or a combination of both. Be creative! Virtually any topic related to the course material is acceptable, provided the project is well-executed.

You may work alone, or in a team of two students. For teams of two, you will be expected to document who did what. Your deliverables will be a **project report** and a **short (15 min) talk** during the final exam slot for this class. If you did experimental work, you will also need to submit your **code**. You will be required to submit a brief (<1 page) **project proposal** by midterm, and to discuss your ideas with me before that date.

Tentative Syllabus (subject to change)

Part I: Mutiple-view Geometry and Reconstruction Algorithms

- **Representation of a three-dimensional moving scene (1.5 hours):** Rigid-body motion, canonical exponential coordinates, Rodrigues formula, Euclidean, affine and projective transformations. (Chapter 2 of Ma et. al.)
- **Image formation (1.5 hours):** Mathematical model for ideal perspective projection and the pinhole camera, other geometric projection models. (Chapter 3 of Ma et. al.)
- **Image primitive and correspondence (1.5 hours):** Photometric features and geometric features, image correspondences and optical flows, feature selection, matching and tracking. (Chapter 4 of Ma et. al.)
- **Two-view geometry (3.0 hours):** Epipolar geometry (discrete and continuous cases), geometric characterization of the essential matrix, and the eight-point, seven-point and six-point algorithms, optimal estimation and reconstruction from two views. (Chapter 5 of Ma et. al.)
- **Camera calibration and self-calibration (1.5 hours):** Camera calibration from a rig, uncalibrated epipolar geometry, the fundamental matrix, camera self-calibration from Kruppa's equations and from projective, affine and Euclidean stratification. (Chapter 6 of Ma et. al.)
- **Multiple-view geometry of points or lines (3 hours):** Two- and three-view geometry review, relation to bi-focal and tri-focal tensors. Multiple-view constraints in terms of matrix rank, reconstruction based on pure point and line features. (Chapter 8 of Ma et. al.)
- **Multiple-view geometry of all incidence relationships (1.5 hours):** A universal rank condition for multiple images of mixtures of points, lines, and planes. (Chapter 9 of Ma et. al.)
- **Multiple-view geometry and symmetry (3 hours):** Symmetric multi-view relations associated with a single view of symmetric structures: reflective, rotational, translational. Symmetry-based 3-D reconstruction. (Chapter 10 of Ma et. al.)

Midterm Project Proposal: A 5 minutes presentation and two-page project proposal from each team.

Part II: Learning for Recognition and Reconstruction

- **Structure from motion and visual SLAM (3.0 hours):** Local feature detection and matching (SIFT, ORB), and bundle adjustment.
- **Reconstruction from regular textures and patterns (3.0 hours):** Transform invariant low-rank texture, camera self-calibration from low-rank texture, reconstructing generalized cylindrical surface, multiple image alignment.
- **Learning to detect vanishing points (1.5 hours):** Conic convolution neural networks for vanishing point detection.
- **Learning to detect wireframes and planes (3 hours):** 2D wireframe detection and parsing, 3D lifting of Manhattan wireframes, plane detection and parsing.
- **Learning to detect and reconstruct symmetric structures (1.5 hours):** Detection and reconstructing objects with reflective, rotational, and translational symmetry.
- **Learning to detect and match global structures in multiple views (3 hours).**
- **Applications in augmented reality and autonomous navigation etc. (3 hours).**

Final Project Presentation and Report: 15 minutes in class presentation for each project. The final report shall be submitted in the format of a conference paper.