CS231A
Computer Vision: From 3D Reconstruction to Recognition

Class Time

M-W; 11:30—13:00PM
CS231A

Instructors

- ssilvio@stanford.edu
- Office: Gates Building, room: 256
- Office hour: tbd and by appointment

- bohg@stanford.edu
- Office: Gates Building, room: 244
- Office hour: Friday 9-10am or by appointment

CAs:
- Andrey Kurenkov, Krishnan Srinivasan, JunYoung Gwak, Yinan Zhang
Lecture 1

Introduction

• An introduction to computer vision
• Course overview
AI is a propelling force of today’s technology
Smart Agriculture

Courtesy of D. Rubin, Stanford

Courtesy of Amazon.com

Courtesy of Agriculture Corner
Health care

Courtesy of D. Rubin, Stanford
Retail

From Imagining the Retail Store of the Future - The New York Times, April 12, 2017
Manufacturing
Transportation and Logistics
Construction Management
Why is this acceleration happening now?
Enabling factors

• Big data

ImageNet, 2009
ShapeNet, 2015
Enabling factors

- Big data
- Faster hardware
Enabling factors

• Big data
• Faster hardware
• New algorithms
  – Representation learning
  – Neural networks
  – Inject learning to deterministic reasoning
1. **Information extraction**: features, 3D structure, motion flows, etc...
2. **Interpretation**: recognize objects, scenes, actions, events
Major areas in Computer Vision

**Space/Geometry**
- Object shape recovery
- Depth estimation
- 3D scene reconstruction

**Semantics/Learning**
- Object detection and pose estimation
- Object tracking
- Scene understanding
Major areas in Computer Vision

**Space/Geometry**
- Object shape recovery
- Depth estimation
- 3D scene reconstruction

**Semantics/Learning**
- Object detection and pose estimation
- Object tracking
- Scene understanding
Recovering 3D models of the environments
Recovering 3D models of the environments
This is critical for autonomous driving or navigation!
Major areas in Computer Vision

**Space/Geometry**
- Object shape recovery
- Depth estimation
- 3D scene reconstruction

**Semantics/Learning**
- Object detection and pose estimation
- Object tracking
- Scene understanding
Detecting and tracking objects in the environments
3D Scene Parsing
Major areas in Computer Vision

**Space/Geometry**
- Object shape recovery
- Depth estimation
- 3D scene reconstruction

**Semantics/Learning**
- Object detection and pose estimation
- Object tracking
- Scene understanding
CS 231A course overview

1. Space/Geometry
   Estimating spatial properties of objects and scene from images through geometrical methods

2. Semantics/Learning
CS 231A course overview

1. Space/Geometry
   Estimating spatial properties of objects and scene from images through geometrical methods

2. Semantics/Learning
   Estimating semantic and dynamic properties of scene elements from images through learning methods
CS 231A course overview

1. Space/Geometry
   Estimating spatial properties of objects and scene from images through geometrical methods

2. Semantics/Learning
   Estimating semantic and dynamic properties of scene elements from images through learning methods
Camera systems
Establish a mapping from 3D to 2D
How to calibrate a camera

Estimate camera parameters such as pose or focal length
Single view metrology

Estimate 3D properties of the world from a single image
Multiple view geometry

Estimate 3D properties of the world from multiple views

Epipolar geometry
Structure from motion

Courtesy of Oxford Visual Geometry Group
Panoramic Photography
3D Modeling of landmarks
Accurate 3D Object Prototyping

Scanning Michelangelo’s “The David”
- The Digital Michelangelo Project
- 2 BILLION polygons, accuracy to .29mm
Augmented Reality

- Magic leap
- Daqri
- Meta
- Etc…
CS 231A course overview

1. Space/Geometry
   Estimating spatial properties of objects and scene from images through geometrical methods

2. Semantics/Learning
   Estimating semantic and dynamic properties of scene elements from images through learning methods
Representations and Representation Learning

Example from Advances in Computer Vision – MIT – 6.869/6.819
Monocular Depth Estimation and Feature Tracking
Optical and Scene Flow

J. J. Gibson, The Ecological Approach to Visual Perception

Lucas-Kanade Feature Tracking over multiple frames. Picture adopted from OpenCV Webpage.

A Database and Evaluation Methodology for Optical Flow. Baker et al. IJCV. 2011

Optimal Estimation for Object Tracking

Wang et al. “Dense Fusion: 6D Object Pose Estimation by Iterative Dense Fusion”, CVPR 2019

Manuel Wühtrich et al. “Probabilistic Object Tracking using a Depth Camera”, IROS 2013
Neural Radiance Fields for View Synthesis
Autonomous navigation and safety

**Mobileye**: Vision systems in high-end BMW, GM, Volvo models. But also, Toyota, Google, Apple, Tesla, Nissan, Ford, etc....

Source: A. Shashua, S. Seitz

Navigation in a Neural Radiance World using a Monocular camera only.

Adamkiewicz, Chen et al. 2021
Personal robotics
More Applications

- Factory inspection
- Assistive technologies
- Surveillance
- Exploration and remote operations
## Syllabus

<table>
<thead>
<tr>
<th>Lecture</th>
<th>Topic</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Introduction</td>
</tr>
<tr>
<td>2</td>
<td>Camera models</td>
</tr>
<tr>
<td>3</td>
<td>Camera calibration</td>
</tr>
<tr>
<td>4</td>
<td>Single view metrology</td>
</tr>
<tr>
<td>5</td>
<td>Epipolar geometry</td>
</tr>
<tr>
<td>6</td>
<td>Multi-view and stereo geometry</td>
</tr>
<tr>
<td>7</td>
<td>Volumetric Stereo</td>
</tr>
<tr>
<td>8</td>
<td>Structure from Motion</td>
</tr>
<tr>
<td>9</td>
<td>Fitting and Matching</td>
</tr>
<tr>
<td>10</td>
<td>Representations and Representation Learning</td>
</tr>
<tr>
<td>11</td>
<td>Depth Estimation, Feature Tracking</td>
</tr>
<tr>
<td>12</td>
<td>Optical and Scene Flow I</td>
</tr>
<tr>
<td>13</td>
<td>Optical and Scene Flow II</td>
</tr>
<tr>
<td>14</td>
<td>Optimal Estimation I</td>
</tr>
<tr>
<td>15</td>
<td>Optimal Estimation II</td>
</tr>
<tr>
<td>16</td>
<td>Neural Radiance Fields I</td>
</tr>
<tr>
<td>17</td>
<td>Neural Radiance Fields II</td>
</tr>
<tr>
<td>18</td>
<td>Guest Lecture</td>
</tr>
</tbody>
</table>

**January**

- Lecture 1: Introduction
- Lecture 2: Camera models
- Lecture 3: Camera calibration
- Lecture 4: Single view metrology
- Lecture 5: Epipolar geometry
- Lecture 6: Multi-view and stereo geometry

**February**

- Lecture 7: Volumetric Stereo
- Lecture 8: Structure from Motion
- Lecture 9: Fitting and Matching
- Lecture 10: Representations and Representation Learning
- Lecture 11: Depth Estimation, Feature Tracking
- Lecture 12: Optical and Scene Flow I
- Lecture 13: Optical and Scene Flow II
- Lecture 14: Optimal Estimation I
- Lecture 15: Optimal Estimation II

**March**

- Lecture 16: Neural Radiance Fields I
- Lecture 17: Neural Radiance Fields II
- Lecture 18: Guest Lecture

Proposal due: February
Mid term: March

**Learning**

- Project presentations

**Geometry**

- Final projects
Prerequisites

• This course requires knowledge of linear algebra, probability, statistics, machine learning and computer vision, as well as decent programming skills.

• Though not an absolute requirement, it is encouraged and preferred that you have at least taken either CS221 or CS229 or CS131A or have equivalent knowledge.

• We will leverage concepts from low-level image processing (CS131A) (e.g., linear filters, edge detectors, corner detectors, etc...) and machine learning (CS229) (e.g., SVM, basic Bayesian inference, clustering, neural networks, etc...) which we won’t cover in this class.

• We will provide links to background material related to CS131A and CS229 (or discuss during TA sessions) so students can refresh or study those topics if needed.
Text books

Required:

Recommended:
- [PB] Probabilistic Robotics, by Thrun, Fox and Burgard, MIT Press, 2005. (PDF of relevant chapter will be provided)
Course assignments

• 1 warm up problem set (HW-0) released today
• 4 problem sets
• 1 mid-term exam
• 1 project

• Look up class schedule for release and due dates.
• Problems will be released through the schedule page and must be submitted through Gradescope (Use code 6P3K6D).
The exam will be on 02/14.
You will be updated with more details, e.g., platform, timing, material to be covered, review sessions etc., as we approach the midterm.
Course Projects

- Replicate an interesting paper
- Comparing different methods to a test bed
- A new approach to an existing problem
- Original research

- Write a 10-page paper summarizing your results
- Release the final code
- Give a final in-class presentation
- SCPD students can send videos instead.

- We will introduce project ideas in 1-2 weeks
- Important dates: look up class schedule
Course Projects

• Form your team:
  – 1-3 people
  – The larger is the team, the more work we expect from the team
  – Be nice to your partner: do you plan to drop the course?

• Evaluation
  – Quality of the project (including writing)
  – Final project in-class presentation (~ TBA minutes spotlight presentations)
Grading policy

- Homeworks: 37%
  - 1% for HW0
  - 9% for HW1, HW2, HW3, HW4 (each)

- Mid term exam: 20%

- Course project: 38%
  - Project proposal 1%
  - Mid term progress report 5%
  - Final report 25%
  - Presentation 7%

- Attendance and class participation: 5%
  - Questions, answers, remarks, Ed posts, Quizzes, OH attendance ...
  - Submitting questions/corrections for new course notes 9-15
  - Class participation are waived for SCPD students. For the project presentation, SCPD students can send videos instead.
Grading policy (HWs)

– 25% will be deducted per day late.
– Four 24-hours one-time late submission “bonuses” are available; that is, you can use this bonus to submit your HW late after at most 24 hours. This is one time deal: After you use all your bonuses, you must adhere to the standard late submission policy.
– Max 2 bonuses can be used per assignment.
– No exceptions will be made.
Grading policy (project)

– If 1 day late, 25% off the grade for the project
– If 2 days late, 50% off the grade for the project
– Zero credits if more than 2 days
– No "late submission bonus" is allowed when submitting your progress report or project report
Are you remote or local during the majority of the winter quarter?

remote

local
What time zone are you in during the majority of the quarter?

<table>
<thead>
<tr>
<th>Pacific</th>
</tr>
</thead>
<tbody>
<tr>
<td>Eastern</td>
</tr>
<tr>
<td>Mountain</td>
</tr>
<tr>
<td>Central</td>
</tr>
<tr>
<td>Europe +Africa</td>
</tr>
<tr>
<td>Asia + Australia</td>
</tr>
</tbody>
</table>
Are you an SCPD student?

yes

no
CS231
Introduction to Computer Vision

Next lecture: Camera systems