3D Vision

Derek Hoiem
University of Illinois

Fall 2021
Today’s class

• A little about me

• Intro to 3D vision

• Course logistics

• 2D-3D Basics
About me

Raised in “upstate” NY
About me

1998-2002
Undergrad at SUNY Buffalo
B.S., EE and CSE

2002-2007
Grad at Carnegie Mellon
Ph.D. in Robotics

2007-2008
Postdoc at Beckman Institute

2009-
Prof in CS at UIUC

2016-
CTO / Chief Scientist
Reconstruct
Computer vision provides situational awareness

What

[Image: Street scene]

Who

[Sun et al 2019]

Where

[Kirillov et al 2019]

[Image: 3D reconstruction]
3D Vision Matters

**Inspection:** Reduce cost and time of inspection to enable frequent inspection and reduce disasters

**Construction:** Reduce schedule cost, risk, and plan deviation to benefit builders, owners, and dwellers

**Driving:** Fewer accidents, less stress

**Robotics:** Do repetitive jobs fast, dangerous jobs safely

Photo credit: https://emerj.com/ai-sector-overviews/how-self-driving-cars-work/

Image: Hu et al. 2019
What is the layout of the environment?

Multiview Reconstruction

Single-view Reconstruction

[Reconstruct] [Zou et al. 2018]
What does the scene look like from new views?

Mesh-based

[Image: Riegler Kolton 2020]

NeRF

[Image: Mildenhall et al. 2020]
Where were the photos taken from?

Structure from Motion (SfM)  Simultaneous Localization and Mapping (SLAM)
How does reality compare to expected?

Alignment, Shape Fitting

[Reconstruct]
What objects are there? What are their poses/shapes?

Semantic Segmentation

Single-view Shape

RGB Image

Predicted Mesh

[Shin et al. 2018]

[Hu et al. 2021]

[Semantic Segmentation]

[Single-view Shape]
My first main research project: single-view 3D reconstruction with Efros, Hebert
Most recent: multi-view 3D Reconstruction

PatchMatch-RL (Lee et al. 2021)
Everything in between

Research
• Robot path planning [IROS 2006]
• Objects in 3D context [CVPR 2006, IJCV 2008, CVPR 2008]
• 3D Object Recognition [CVPR 2007, ECCV 2010, CVPR 2018]
• 3D Photo Manipulation [SG 2007, SGA 2011]
• Occlusion Boundaries [ICCV 2007, IJCV 2011]
• RGBD Scene Analysis [ECCV 2012, IJCV 2019]
• Object 3D shape estimation [CVPR 2013, CVPR 2015, ICCV 2017]
• 3D material recognition [CVPR 2016]

Commercial Application
• Reconstruct: SfM, SLAM, MVS, meshing, recognition, registration

But I still have a lot to learn!
This Class

• Learn fundamentals of 3D vision
  – Lectures on Thursdays

• Learn state-of-the-art
  – Discuss papers you select and read on Tuesdays

• Improve research skills
  – Identify potential directions: survey, paper reports
  – Design proof-of-concept: research proposal
  – Perform PoC, re-assess: research paper
Prerequisites

• Graduate-level computer vision (CS 543 or equivalent)

• Engaged or interested in 3D Vision research
Materials

- Website: https://courses.engr.illinois.edu/cs598dwh/fa2021/
  - Syllabus
  - Schedule
  - Paper selection/reports
Paper Readings

For each topic

[Before Thursday class]
1. **Group assignment.** Groups are (randomly) assigned by the professor and listed in Paper Selection. One tab for each topic, one row per group.

[ Before Tuesday class]
2. **Scribe.** Group selects a scribe. Whoever has been scribe fewest times should be scribe next. In case of tie, can choose by interest.
3. **Paper selection.** The scribe chooses a topical paper in consultation with the other group members by end of day Thursday and puts title/link next to group in Paper Selection. No two groups can choose the same paper! First to claim the paper gets it.
4. **Paper reading and review.** By 10:45am Tuesday, each group member (including scribe) submits their reviews using the Review Form.

[In class Tuesday]
5. **Discussion.** In class, students split into groups and discuss the ideas of the paper and ideas for future work or other applications.
6. **Summary.** During discussion period, scribe consolidates discussion in one summary slide. Copy-paste the template under the topic and fill in the slide. Can include figures from paper. Put slides in group order.
7. **Report out.** Scribe presents summary to class.
Course Project

1. Survey
   • Assigned group
   • Choose different topic for each group
   • 4-6 page report: overview, taxonomy, evaluation, analysis, research ideas

2. Research Proposal
   • You form group
   • Choose research proposal idea
   • 2-3 page report: motivation, related work, proposed approach, contributions, significance, planned experiments including proof-of-concept

3. Project Report
   • Same group as proposal
   • Perform proof-of-concept experiments
   • 4 page report: intro, approach, PoC results, recommendations

Reviews: everyone reviews one survey and one proposal
Grading

• Paper reviews and discussion: 50%
  – Must do at least 10 for full points
  – ½ credit if review is unsatisfactory or discussion is missed

• Course project: 50%
  – Survey 15%
  – Proposal 15%
  – Report 15%
  – Reviews 5%
  – Grading is “satisfactory” (full credit), “needs improvement” (3/4 credit), “unsatisfactory” (1/2 credit); can be resubmitted once if necessary

• Late policy
  – no credit for late reviews
  – project component penalty is 1% of course total per day
Academic Integrity

• All work you submit should be your own – do not copy any text from any online reviews or papers
  – Cite sources diligently

• If your research project builds on prior/ongoing work, discuss with professor first

• Violations will be penalized through official channels
COVID-19 Policy

• Students who feel ill must not come to class. In addition, students who test positive for COVID-19 or have had an exposure that requires testing and/or quarantine must not attend class.
  – You will not lose review/discussion points for this

• All students, faculty, staff, and visitors are required to wear face coverings in classrooms and university spaces.
Getting help outside of class

Office hours
• For help with projects or papers or other complex questions, see professor after class or another arranged time

Slack:
• For discussion within student groups or logistical questions
https://join.slack.com/t/3dvision-fa21/shared_invite/zt-u1yy4vk1-8oEBalkCVT15GhQeoLaF7g

Readings/Textbook
• See webpage
Questions about class structure/content?
Basics of Cameras: **What is a pixel?**

**Image coordinate**

**CCD cell**

**3D ray**

$u, v$

$P = \begin{bmatrix} X \\ Y \\ Z \end{bmatrix}$
How do we map from 3D to 2D?

\[
p = K P
\]

\[
p = \begin{bmatrix} u \\ v \end{bmatrix} = \begin{bmatrix} f & 0 & u_0 \\ 0 & f & v_0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix}
\]
Homogeneous coordinates

Homogeneous Coordinates

\[
\begin{bmatrix}
  x \\
  y \\
  k
\end{bmatrix}
= \begin{bmatrix}
  w_x \\
  w_y \\
  w_k
\end{bmatrix}
\Rightarrow
\begin{bmatrix}
  \frac{w_x}{w_k} \\
  \frac{w_y}{w_k}
\end{bmatrix}
= \begin{bmatrix}
  \frac{x}{k} \\
  \frac{y}{k}
\end{bmatrix}
\]

= a Ray

= a Point
Basic geometry in homogeneous coordinates

• Line equation: \( ax + by + c = 0 \)

\[
\begin{bmatrix}
  a_i \\
  b_i \\
  c_i \\
\end{bmatrix}
\]

\( line_i \)

• Append 1 to pixel coordinate to get homogeneous coordinate

\[
\begin{bmatrix}
  u_i \\
  v_i \\
  1
\end{bmatrix}
\]

\( p_i \)

• Line given by cross product of two points

\[
line_{ij} = p_i \times p_j
\]

• Intersection of two lines given by cross product of the lines

\[
q_{ij} = line_i \times line_j
\]
How do we map from 2D to 3D?

Sometimes called a “bearing”

$$K^{-1}p = wP$$

$$\begin{bmatrix}
    \frac{1}{f} & 0 & \frac{-u_0}{f} \\
    0 & \frac{1}{f} & \frac{-v_0}{f} \\
    0 & 0 & 1
\end{bmatrix}
\begin{bmatrix}
    u \\
    v \\
    1
\end{bmatrix}
= w
\begin{bmatrix}
    x \\
    y \\
    z
\end{bmatrix}$$
Rotation and translation map from “world” coordinates to “camera” coordinates

\[
X_c = [R \ t] X_w \quad \text{x} = K [R \ t] X_w
\]

- \( X_c \): Image Coordinates: \((u, v, 1)\)
- \( X_w \): World Coordinates: \((X, Y, Z, 1)\)
- \( K \): Intrinsic Matrix (3x3)
- \( R \): Rotation (3x3)
- \( t \): Translation (3x1)
Properties of 3D rotation matrix

\[ R = \begin{bmatrix}
    r_{11} & r_{12} & r_{13} \\
    r_{21} & r_{22} & r_{23} \\
    r_{31} & r_{32} & r_{33}
\end{bmatrix} \]

\[ R^{-1} = R^T \]

\( R \) is orthonormal:

\[ R = \begin{bmatrix}
    r_1 \\
    r_2 \\
    r_3
\end{bmatrix} \quad \|r_i\| = 1 \quad r_i^T r_j = 0 \quad \|RX\| = \|X\| \]
Rotation matrix sudoku

• Solve for missing $r$ values (up to sign ambiguity)

\[
\mathbf{R} = \begin{bmatrix}
  r_{11} & r_{12} & ? \\
  r_{21} & r_{22} & ? \\
  r_{31} & ? & ? \\
\end{bmatrix}
\]

\[
\mathbf{R} = \begin{bmatrix}
  r_{12} & ? \\
  r_{21} & r_{22} & ? \\
  ? & ? & ? \\
\end{bmatrix}
\]
Questions to consider

1. What is the camera’s position in world coordinates, given $\mathbf{R}$ and $\mathbf{t}$?

2. What additional information can enable recovering a 3D geometry coordinate from a 2D pixel coordinate?

3. Suppose a camera images a star (~infinite distance point). If the camera translates without rotating, what is the effect on the pixel position of the star?
Final comments

• To do
  – Review web page and syllabus
  – Start planning with your group which paper to do for next Tuesday

• Next class: two-view stereo

• Questions?