**MISC**

- Instructor: Qixing Huang (UT Austin)
- Time: Beijing Time 10:00 am
- Send me emails for questions ([huangqx@cs.utexas.edu](mailto:huangqx@cs.utexas.edu))
- Webpage: https://www.cs.utexas.edu/~huangqx/
Introduction
Geometry Processing Pipeline

Reconstruction
Processing & Analysis
Printing
3D Vision

Recovering the Underlying 3D structures from Images

Structure-from-motion
Multi-view Stereo
Pose Estimation

RGB images
RGB-D images
Large-scale online repositories

- **3D Warehouse**: 3M models in more than 4K categories
- **Yobi3D**: 50k Indoor Scenes
- **3DFront**
3D Vision (2015-)

RGB (RGBD) Images → 3D Structures → Understanding (Synthesis)

- Pose Estimation
- Structure-from-motion
- Multi-view Stereo
- Classification
- Segmentation
- Detection
2D Vision versus 3D Vision

2D Vision
- Classification

3D Vision
- Segmentation
- Detection
Data Vectorized Form Representation ML/Optimization Algorithms Vectorized Form
3D Representations

- Volumetric
- Triangular mesh
- Point cloud

- Multi-view
- Scene-graph
- Semantic segments
Applications
3D mapping

Semantic 3D model of Enschede, Netherlands

Semantic 3D model of Dortmund, Germany

Semantic 3D model of Zurich, Switzerland

Semantic 3D model of a synthetic city [Cabezas et al. 2015]

https://prs.igp.ethz.ch/research/current_projects/large-scale_reconstruction.html
Performance capture

https://www.engadget.com/2015/03/08/using-the-kinect-for-motion-capture/
Robotics

[Lenz et al. 15]
Autonomous driving

[Yurtsever et al. 19]

Reverse engineering
Interdisciplinary field

- Graphics
- Robotics
- Computer Vision
- Biology
- Architecture
- Medicine
- Geometry Processing & 3D Vision
Topics to be covered
Topic I: 3D Reconstruction

- Scanning
- Scan registration
- Surface reconstruction
- Structure-from-motion
- Multi-view stereo
- Map synchronization
Topic I: 3D Reconstruction

• Scanning

• Scan registration

• Surface reconstruction

• Structure-from-motion

• Multi-view stereo

• Map synchronization
Topic II: How to represent 3D Data

- Triangular mesh
- Part-based models
- Implicit surface
- Light Field Representation
- Point cloud
Conversion between different representations

- Implicit -> mesh (Marching Cube)
Conversion between different representations

- Pointcloud $\rightarrow$ Implicit $\rightarrow$ Mesh

[Kazhdan et al. 06]
Two recommended books

1. Point-Based Graphics
2. Polygon Mesh Processing
Topic III: How to understand 3D Data

• Design algorithms to extract semantic information from one or a collection of shapes

[van Kaick et al. 11]

Matching

Retrieval

[Karz and Tal 03]

Segmentation

[Mitra et al. 06]

Classification & Clustering

[Funkhouser et al. 05]
We focus on the basics
Recent trends in AI

• Deep reinforcement learning

• Robot learning
  – Manipulation

• Autonomous driving

• Visual navigation

• Fairness in machine learning
Building blocks

• Localization
  – Registration

• Recognition under different representations

• Matching
  – Linking objects from different domains
Mathematics is important
Geometry

- Linear algebra
- Perspective geometry
- Differential geometry/topology
Numerical Optimization is Important
Optimization problems

• We will see time and time again:

\[ P : \min_{x \in D} f(x) \]

Translate into Optimization problem

Real problem
Examples in Vision/Robotics/NLP/ML/Graphics/

Example of the contrary?

This course: how to formulate P, how to solve P, and what are the guarantees

Many similar domains: Graphics/Robotics/Vision
Why bother how to solve P and what are the guarantees

- There are plenty of optimization softwares

  **Solvers:** CPLEX, Mosek, Gurobi, ECOS, Clp, Knitro, Ipopt,...

  **Modeling languages:** YALMIP, CVX, GAMS, AMPL, JuMP,...

- Almost all algorithms are data-dependent and can perform better or worse on different problems and data sets

- In many cases, studying P leads to new algorithms and... research papers
Categories of optimization models

• Linear vs. Nonlinear

• Convex vs. Nonconvex

• Continuous vs. Discrete

• Deterministic vs. Stochastic

*We see all of them in various 3D Vision tasks*
Example 1: Bundle adjustment

\[
\psi_f = \sum_{(i,j) \in S} \Psi(\|F_{ij}\|) = \sum_{(i,j) \in S} \frac{1}{2} \|F_{ij}(x_i, \phi_j, t_j, y_j)\|^2
\]

3D locations

Camera parameters
Example 2: Surface fitting

• Important in reverse engineering

[Flory and Hofer’ 10]

A surface fitting example. (Left) The initial setup. (Right) The unsigned distance between point cloud (grey) and surface (blue) is iteratively minimized. The approximating surface is deformed until it fits the point cloud.
Example 3: MRF Inference

\[
\arg \min_{w_1 \ldots w_N} \sum_{n=1}^{N} U_n(w_n) + \sum_{(m,n) \in C} P_{mn}(w_m, w_n),
\]

**Unary terms**
(compatibility of data with label \( w \))

**Pairwise terms**
(compatibility of neighboring labels)

The literature reflects almost all advances in optimization during the past decade:
Trends in optimization

- Non-linear optimization
- Convex optimization
- Non-convex optimization
Trends in optimization

Second-order methods

First-order methods

Distributed optimization
Homework
Assignment 1: dense reconstruction

Multi-view stereo
[Furukawa and Ponce]

Geometry reconstruction
Assignment 2: primitive extraction

Slippage analysis
[Gelfand and Guibas’ 04]

Deep Primitive
[Li et al. 04]
Optional final project
Potential project topic 1

• Image-based modeling from internet images
  – A few images
  – Non-identical objects
  – Symmetries
  – Relative pose
  – Machine learning
Potential project topic 2

• Single-view reconstruction from real images
  – 3D Representation?
    • Volumetric/point cloud/part-based
  – Network architecture
  – Domain adaptation
  – Depth/semantic labels
  – Human
  – Hand

[Song et al. 17]

Output: semantic scene completion
Potential project topic 3

- Neural networks for 3D representations
  - Implicit surface
  - Multi-view
  - Point cloud
  - Mesh?
  - Scene graph?
  - Physics?
  - Hybrid?

[Wu et al. 16]
Potential project topic 4

- Geometry understanding
  - Task
    - Normal/Curvature
    - Feature extraction
    - Segmentation
    - Part/object decomposition
    - Correspondence
    - Affordance
  - Data
  - Thousands of categories
Potential project topic 5

- Reconstruction from hybrid sensors
  - Low-res depth scan + High-res stereo
  - Depth + shading
  - Web cameras + internet images
  - 360 images + internet images
  - Street views + images from drones
Potential project topic 6

• Robot 3D Vision
  – Salient views
  – Next-best-view
  – Robot grasping
  – Human-object interaction
  – Model-based View planning
  – Active vision
  – Policies for exploration
Potential project topic 7

- Structure recovery via optimization
  - 2D human pose -> 3D human pose
  - Structure from symmetry
  - Structure from template
  - Shape from shading

- Local/global convergence
- Exact recovery condition

[Zhou et al. 16]
Potential project topic 8

• Map synchronization
  – Multiview structure from motion
    • Pose
    • Pose + Point cloud
  – Geometric alignment of point clouds
    • Pose
    • Pose + shape
  – Global/Local convergence
  – Exact recovery condition
Potential project topic 9

• Uncertainties
  – Human pose estimation
  – Depth prediction
  – Camera poses in MVSFM
  – Geometry reconstruction
  – Geometry understanding
  – Quantification/Visualization
Scanning
Geometry Reconstruction Pipeline
Depth Sensing

- Range acquisition
  - Contact
    - Mechanical (CMM, jointed arm)
    - Inertial (gyroscope, accelerometer)
    - Ultrasonic trackers
    - Magnetic trackers
  - Transmissive
    - Industrial CT
    - Ultrasound
    - MRI
  - Reflective
    - Non-optical
      - Radar
      - Sonar
    - Optical
Touch Probes

- Jointed arms with angular encoders
- Return position, orientation of tip

Faro Arm – Faro Technologies, Inc.
Depth Sensing

Optical methods

Passive

Shape from X:
- stereo
- motion
- shading
- texture
- focus
- defocus

Active

Active variants of passive methods
- Stereo w. projected texture
- Active depth from defocus
- Photometric stereo

Time of flight

Triangulation
Pulsed Time of Flight

- Basic idea: send out pulse of light (usually laser), time how long it takes to return

\[ d = \frac{1}{2} c \Delta t \]
Triangulation
Depth from Stereo Uses Parallax

- The Kinect analyzes the shift of the speckle pattern by projecting from one location and observing from another.
Triangulation: Moving the Camera and Illumination

• Moving independently leads to problems with focus, resolution

• Most scanners mount camera and light source rigidly, move them as a unit
Triangulation: Moving the Camera and Illumination
Triangulation: Extending to 3D

- Possibility #1: add another mirror (flying spot)
- Possibility #2: project a stripe, not a dot
Pattern Design
Structured Light General Principle

Project a known pattern onto the scene and infer depth from the deformation of that pattern.

Zhang et al, 3DPVT (2002)
Time-Coded Light Patterns

- Assign each stripe a unique illumination code over time [Posdamer82]
Time-Coded Light Patterns

• Assign each stripe a unique illumination code over time [Posdamer82]
Binary Coding Example

Codeword of this pixel: 1010010 -> Identifies the corresponding pattern stripe
Temporal vs. Spatial Continuity

• Structured-light systems make certain assumptions about the scene

• Temporal continuity assumption
  – Assume scene is static
  – Assign stripes a code over time

• Spatial continuity assumption
  – Assume scene is one object
  – Project a grid, pattern of dots, etc.
Grid Methods

• Assume exactly one continuous surface

• Count dots or grid lines

• Occlusions cause problems

• Some methods use dynamic programming
Codes Assuming Local Spatial Continuity

Codeword
Codes Assuming Local Spatial Continuity

[Zhang et al. 02]

[Ozturk et al. 03]
Kinect combines structured light with two classic computer vision techniques: depth from focus, and depth from stereo.
The Kinect uses infrared laser light, with a speckle pattern

Shpunt et al, PrimeSense patent application
US 2008/0106746
Next lecture

- Registration