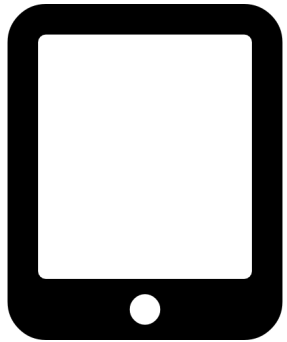
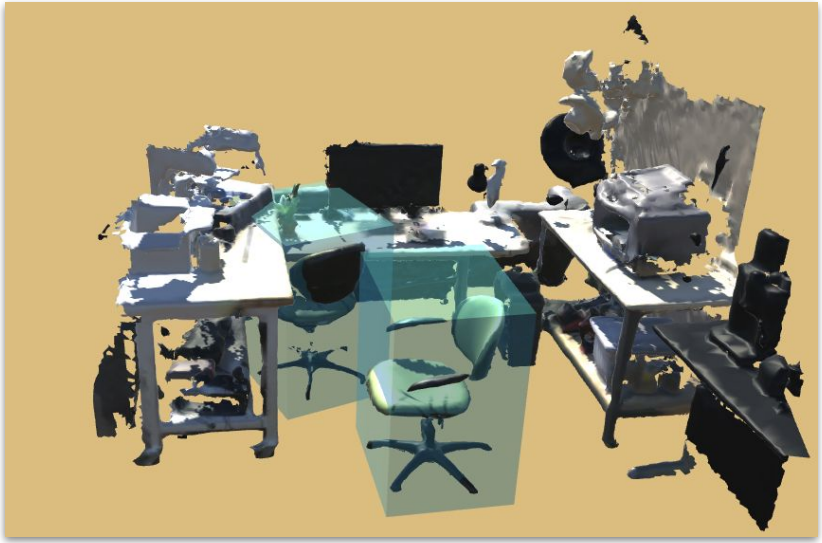
Augmented Annotations

Final Presentation



James Lin, Vedant Saran

Augmented Annotations (AA) is a 3D data capture utility on iOS. Users scan their environment while simultaneously annotating objects with labeled bounding boxes.



Introduction

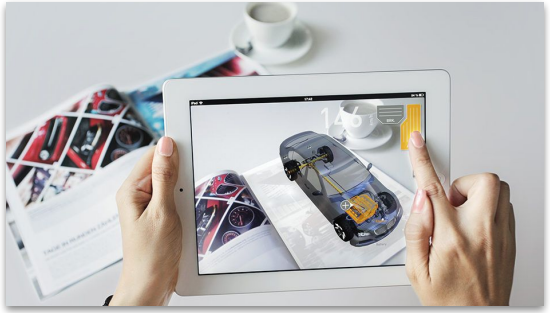
Motivation

- 3D data annotation is a slow, tedious process
- Efficient solutions require infrastructure
  - Mechanical Turk
  - Undergrads
- Huge, unspoken time sink in research
- Expensive to make custom datasets

Dataset	Bounding Boxes
SUN RGB-D	64,595
KITTI	80,256

Key insight

- Realtime SLAM accuracy >= human accuracy
- Scanning has been commodified
  - Cheaper, portable sensors
  - Augmented reality (AR)



AR application mock-up

Intel Realsense D435

Our solution

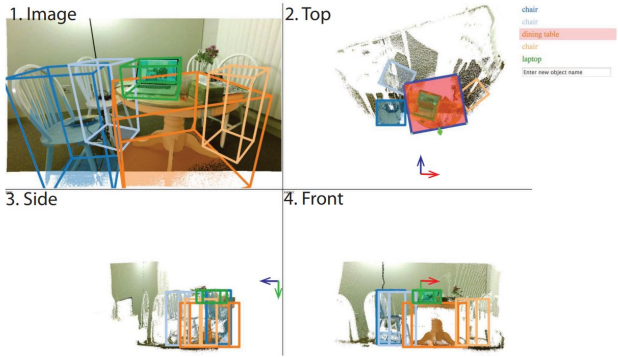
- Consolidate **scanning** and **annotation** processes
- Platform: iOS device (iPad)
- Scan with Occipital’s Structure Sensor
- Add/edit bounding boxes as AR objects

Structure sensor mounted on an iPad

Related Works

Traditional methods

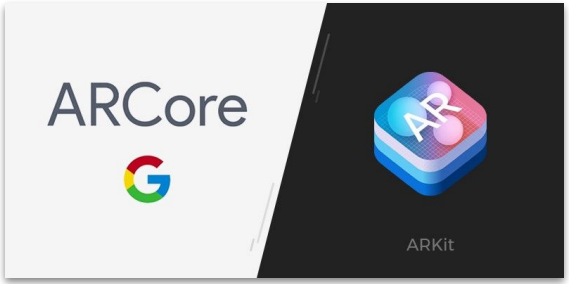
- Digital annotation software + outsourced labor
  - Mechanical Turk, oDesk
- Downsides
  - Infrastructure requirement
  - User training requirement
  - Quality-assurance difficult



Screenshots from SUN RGB-D's annotation tool

Augmented reality

- Main workflow: placing localized objects in the world
- Heavy reliance on real-time SLAM
  - Scans are immediately discarded



Popular mobile AR libraries

Example of localized object placement

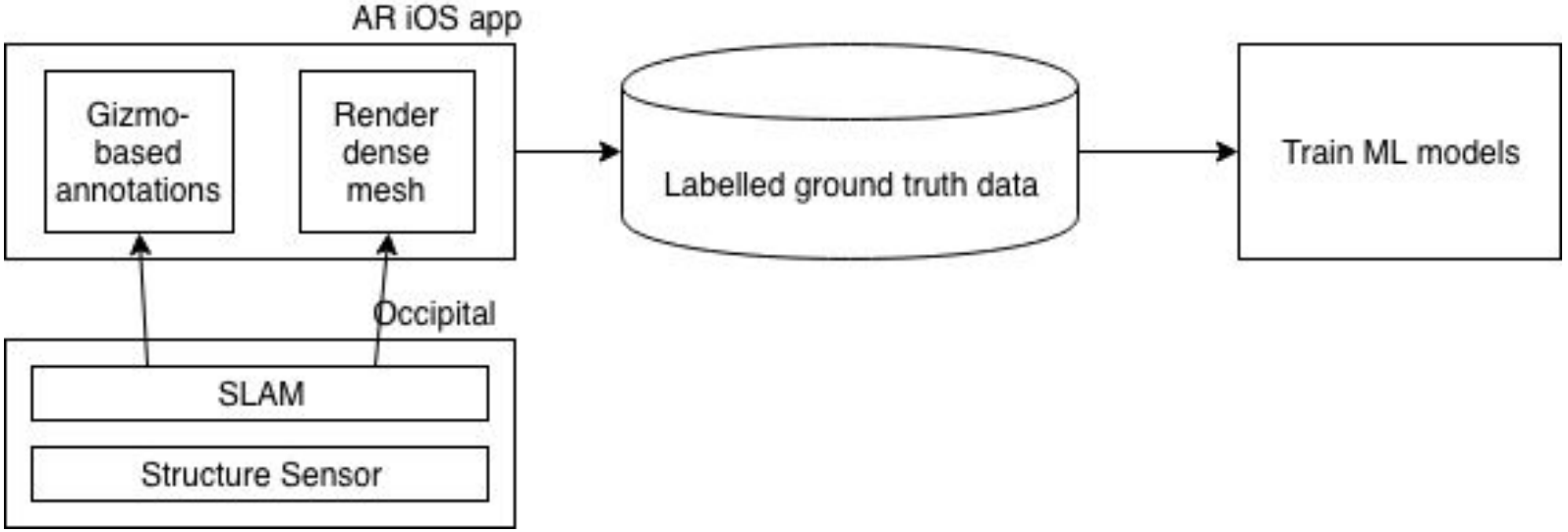
Augmented reality meets computer vision

Real scene augmented with synthetic cars (Ours)

Popular mobile AR libraries

Methods

Architecture



Scanning

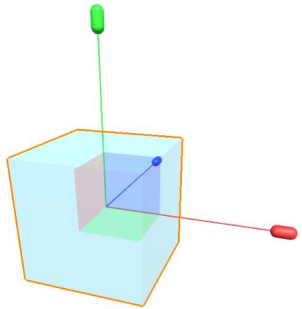
Occipital Structure Sensor

- Developed by Occipital Inc. in collaboration with Prime Sense

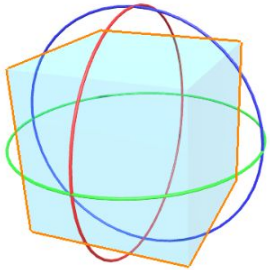
Parameter or specifications	Value or description
Length × Width × Height	119.2 mm × 27.9 mm × 29.0 mm
Min/Max sensing distance	40/350 cm
Resolution for depth data	VGA (640 × 480) QVGA (320 × 240)
Frame rate	30/60 fps
Power source	USB charged battery with 3–4 h of active sensing
Data transfer	Wi-Fi/USB
Field of view	58° Horizontal and 45° Vertical standard lens
System on a chip (SOC)	Prime Sense chip (Heindl 2014)

Annotation: 3D bounding boxes

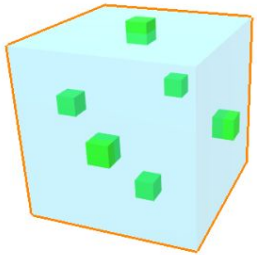
- **Gizmos**: touch-and-drag control modules
- Based off of industry 3D manipulation tools
  - Unity, Maya, etc.



Position tool



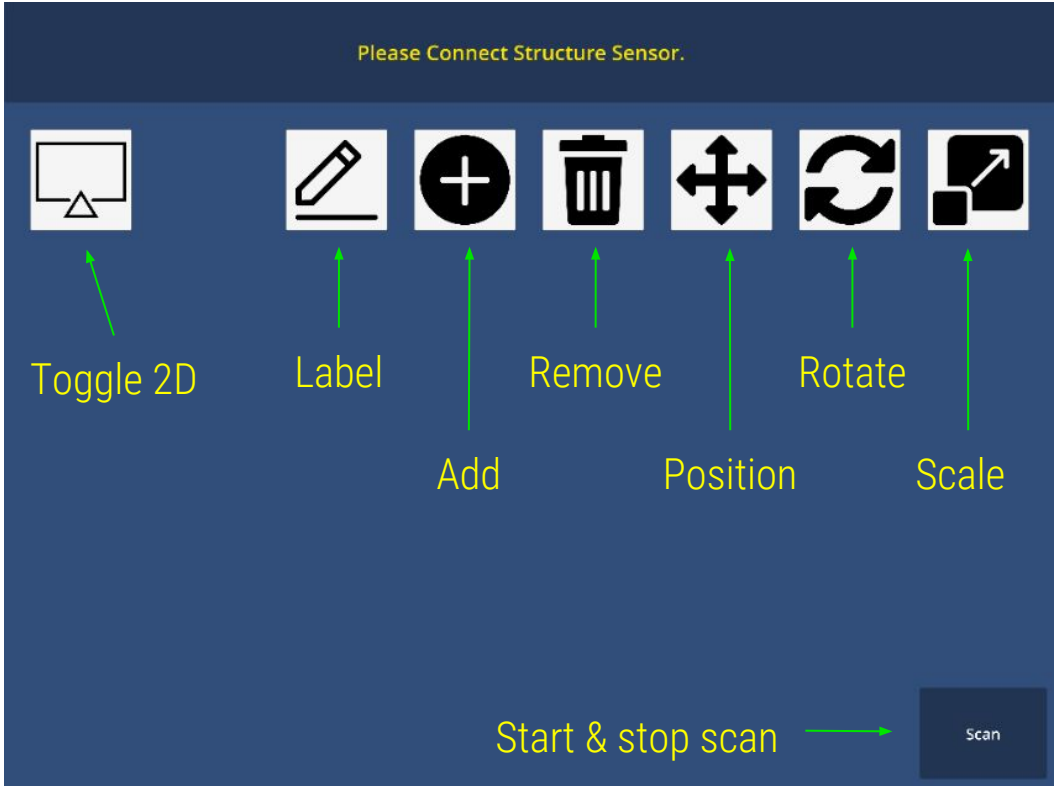
Rotate tool



Scale tool

Functionality
Add
Remove
Position
Rotate
Scale
Label

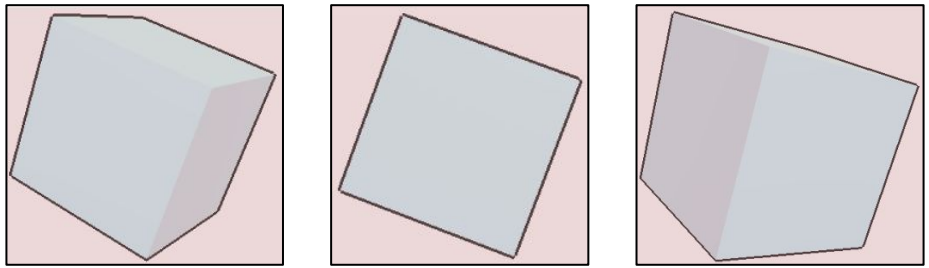
Annotation: user interface



Annotation: example usage

Annotation: 2D bounding boxes

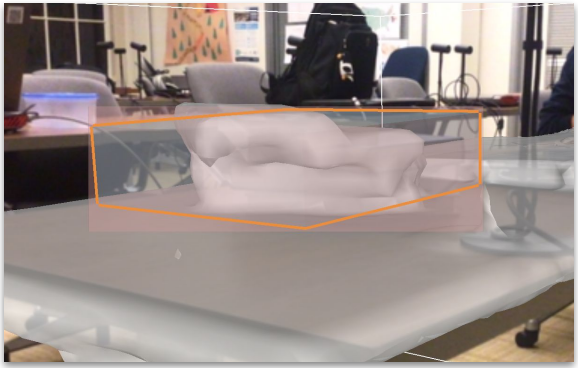
- Can project 3D ⟶ 2D bounding box for any perspective
  - **Pro**: High convenience / speed of producing data
  - **Con**: 2D bounding boxes larger than necessary

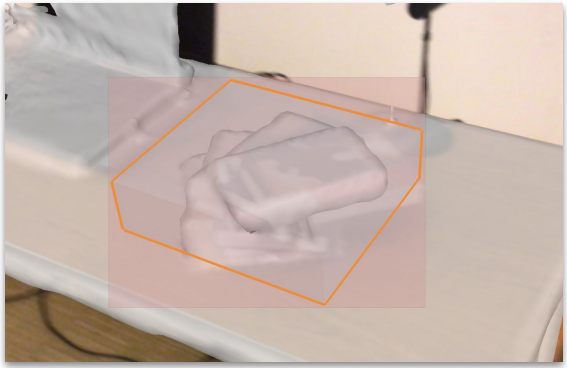


2D bounding boxes at different perspectives

Annotation: 2D bounding boxes

- Different perspectives ⟶ different 2D bounds

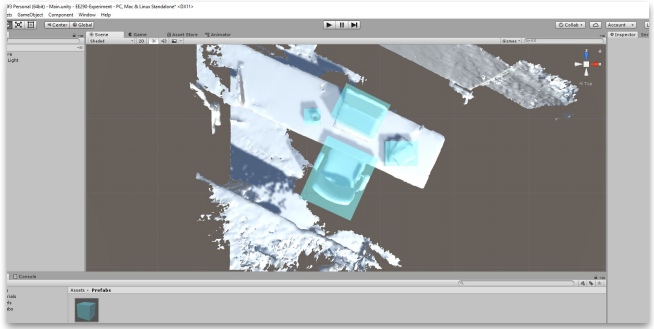




Results

Comparison with SUN RGB-D

- Conducted trials comparing SUN RGB-D's method to our app
- Recorded time to completion
- Task: create and label bounding boxes over 4 objects

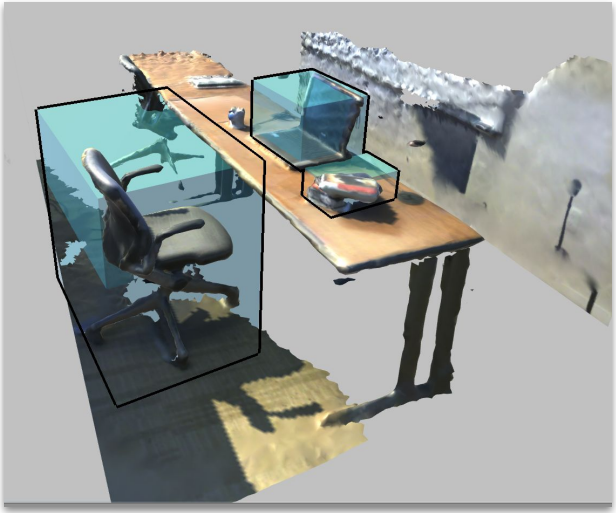


Unity editor, used for desktop trials

Physical setup for scanning

User feedback

- Pros
  - Intuitive task and control scheme
  - Easy to view boxes from different perspectives
- Cons
  - Require lots of physical movement
  - Gizmo controls occasionally unwieldy/unstable



Example output from our app

Comparison with SUN RGB-D

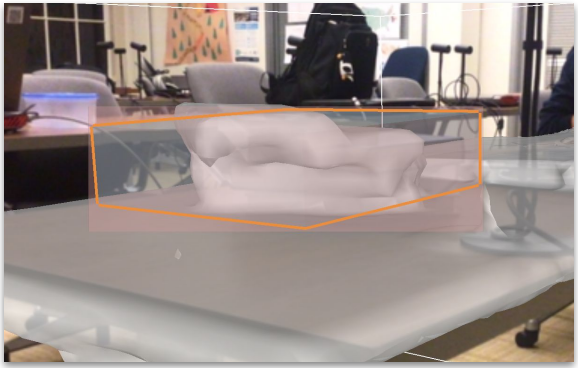
Trial	1	2	3	Average
SUN RGB-D S	0:46	0:46	0:46	5:13
SUN RGB-D A	4:20	4:04	4:57	
AA S + A	2:58	3:23	4:05	3:28

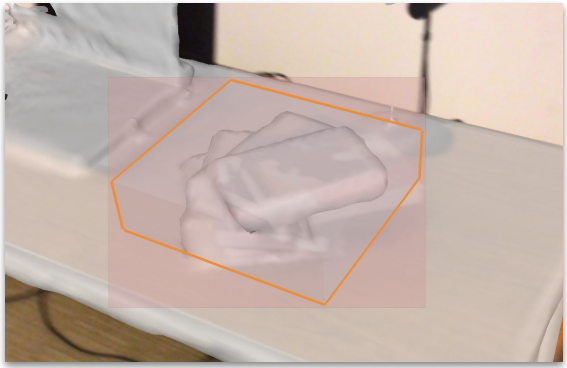
S: Scanning
A: Annotation

33.5% reduction in time!

Annotation: 2D bounding boxes

- Dramatically faster for 2D bounding boxes!





Conclusion

Summary

- Developed an application to combine the scanning and annotation processes of 3D data collection
- Conducted user studies on the effectiveness of our method
- Shown to be faster than desktop-based annotation systems

Future work

- Intelligent bounding box placement
  - Applying corrections faster than starting from scratch
- Smoothing out user experience
- Applying same strategy to other annotation types
  - Semantic segmentation