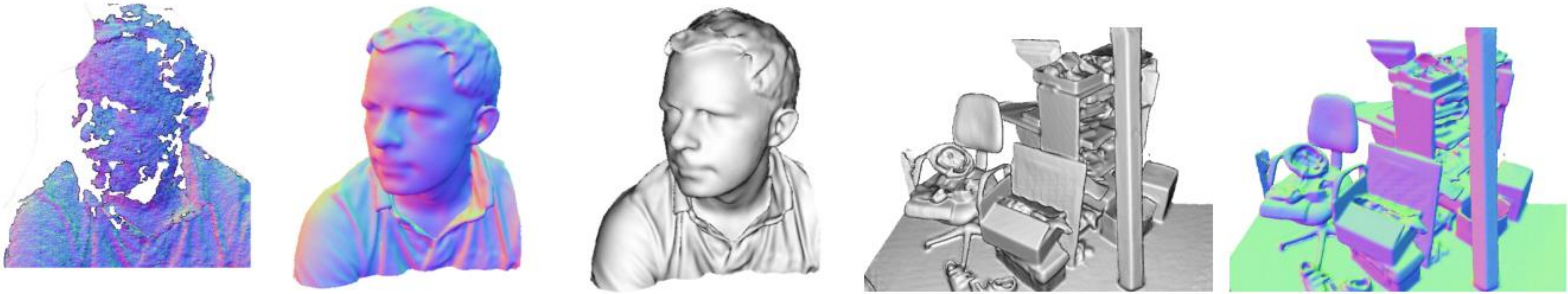
KinectFusion

Lecturer: Zhi-Hao Lin

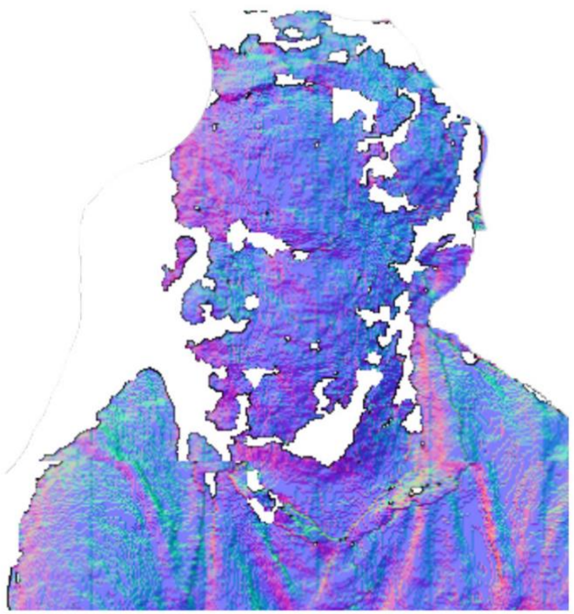
Date	Topic	Papers	Lecturer 1	Lecturer 2	Archeologist 1	Archeologist 2	Private Investigator	Industrial Practitioner	Critic	Graduate Student	Hacker 1	Hacker 2	Hacker 3	Discussion
Sept 12	MVS	KinectFusion	Zhi-Hao		Shenlong		Shenlong	Albert Zhai	Zhi-Hao	Albert Zhai	ICP (Zhi-Hao)	2D Depth Fusion (Shenlong)	-	
Sept 17	SLAM	DROID-SLAM	rmoan2		ying22	xialinh2	jw116	jrc9	zhiyul6	yuezeng4	ksa5	wenqij5		
Sept 19	Learning-based SFM	Camera as Rays	dyyao2		ads10	haoyuyh3	yuqunwu2	jiahuad2	amitabh3	haoz19	amballa2	ozgurk2		
Sept 26	Mesh & Tedrahedron	DMTet	zixuan32	ying22	jw116	wch7	hongchix	runpeid2	hanlinm2	jiaweiz9	haozhe3	weijiel4		
Oct 8	Neural Volumes	DeepSDF	xialinh2	ads10	junzhew3	zixuan32	ziyang8	mkg7	rmoan2	vlasz2	junkun3	jackiel4		
Oct 10	Deep Learning on 3D	fVDB	haoyuyh3		ksa5	wenqij5	dyyao2	hongyig3	haozhen3	haozhe3	tcheng12	shaowei3		
Oct 17	Differentiable Rendering	NVDiffRast	wch7	junzhew3	amballa2	ozgurk2	guangy2	meih3	nisar2	junkun3	cconway4	yufengl2		
Oct 22	Neural Radiance	Gaussian Splats	nisar2	ziyang8	haozhe3	weijiel4	yuezeng4	xialinh2	ying22	tcheng12	jrc9	zhiyul6		
Oct 24	Neural Surface	NeuS	wenqij5	haozhen3	junkun3	jackiel4	haoz19	haoyuyh3	ads10	cconway4	jiahuad2	amitabh3	yuqunwu2	
Oct 29	Neural Inverse rendering	Rendering synthetic object	ozgurk2	guangy2	tcheng12	shaowei3	jiaweiz9	wch7	jw116	jrc9	runpeid2	hanlinm2	hongchix	
Oct 31	Monocular Geometry	Recovering surface layout f	weijiel4	yuezeng4	cconway4	yufengl2	vlasz2	zixuan32	junzhew3	jiahuad2	mkg7	rmoan2	ziyang8	
Nov 5	Few-View Geometry	Mast3R	jackiel4	haoz19	jrc9	zhiyul6	ksa5	wenqij5		runpeid2	hongyig3	haozhen3	dyyao2	
Nov 7	3D Editing	Enhancing Photorealism Er	shaowei3	jiaweiz9	jiahuad2	amitabh3	amballa2	ozgurk2	yuqunwu2	mkg7	meih3	nisar2	guangy2	
Nov 19	3D Simulation	PhysGaussian	yufengl2	vlasz2	runpeid2	hanlinm2	haozhe3	weijiel4	hongchix	hongyig3	xialinh2	ying22	yuezeng4	
Nov 21	3D Generation	Get3D	zhiyul6		mkg7	rmoan2	junkun3	jackiel4	ziyang8	meih3	haoyuyh3	ads10	haoz19	
Dec 3	Parametric Articulated Shapes	SMPL	amitabh3	yuqunwu2	hongyig3	haozhen3	tcheng12	shaowei3	dyyao2	ksa5	wch7	jw116	jiaweiz9	
Dec 5	Dynamic 3D Understanding	Dynamic Fusion	hanlinm2	hongchix	meih3	nisar2	cconway4	yufengl2	guangy2	amballa2	zixuan32	junzhew3	vlasz2	
	Time		12		12		6	6	6	6		12		10

KinectFusion: Real-Time Dense Surface Mapping and Tracking*

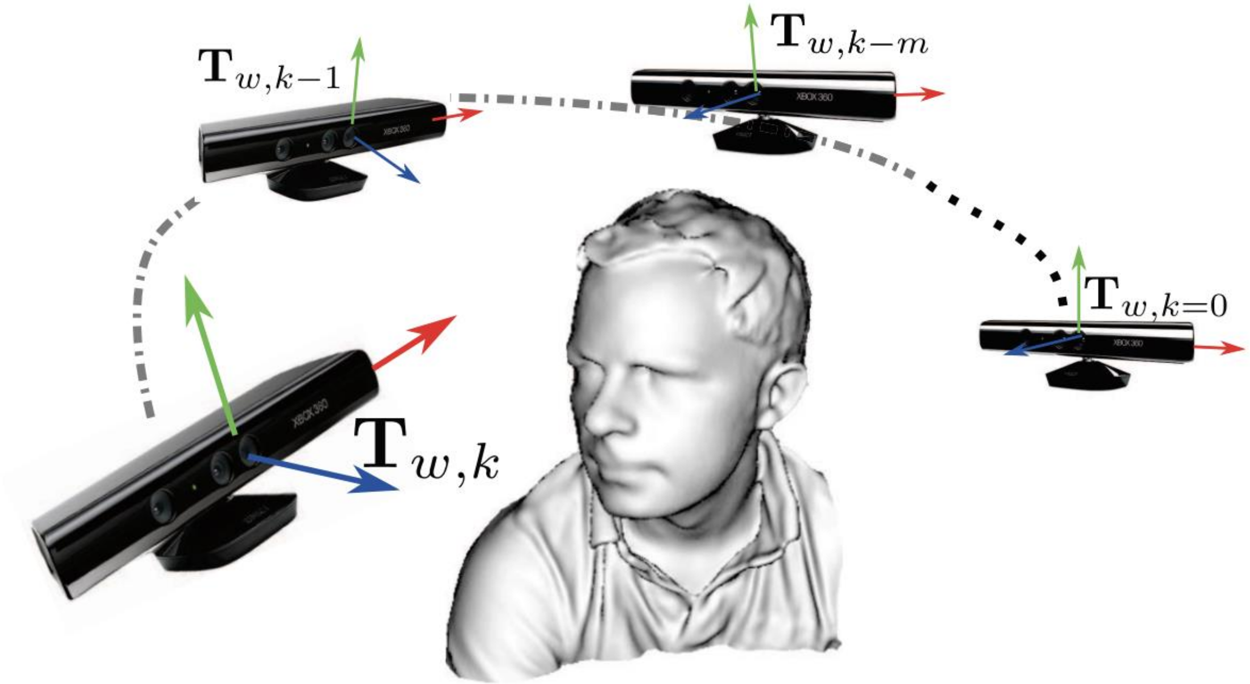
Richard A. Newcombe, Imperial College London
Shahram Izadi, Microsoft Research
Otmar Hilliges, Microsoft Research
David Molyneaux, Microsoft Research, Lancaster University
David Kim, Microsoft Research, Newcastle University
Andrew J. Davison, Imperial College London
Pushmeet Kohli, Microsoft Research
Jamie Shotton, Microsoft Research
Steve Hodges, Microsoft Research
Andrew Fitzgibbon, Microsoft Research



KinectFusion: Background



Depth Sensor Sequences



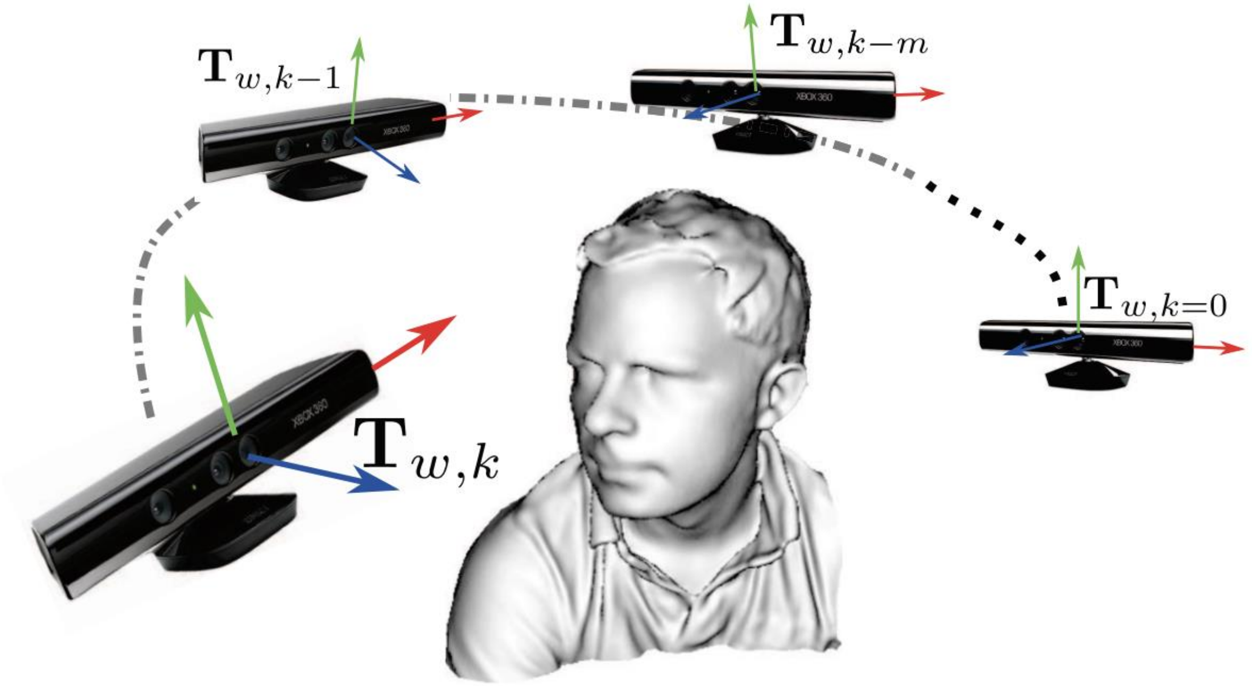
Complete Shape and Camera Poses

Image credit: Newcor

KinectFusion: Background

Tracking: Estimate camera pose T

Mapping: 3D reconstruction



Complete Shape and Camera Poses

Image credit: Newcor

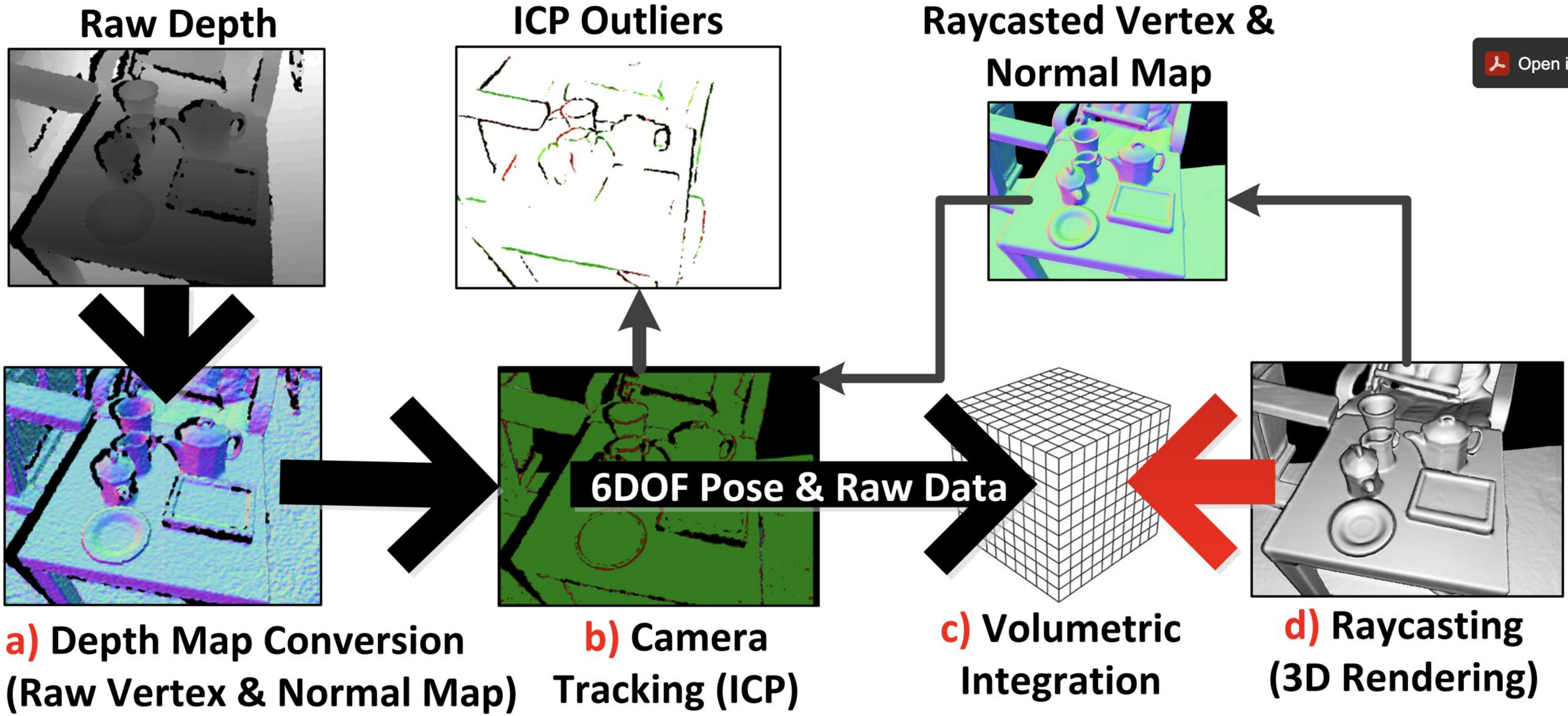
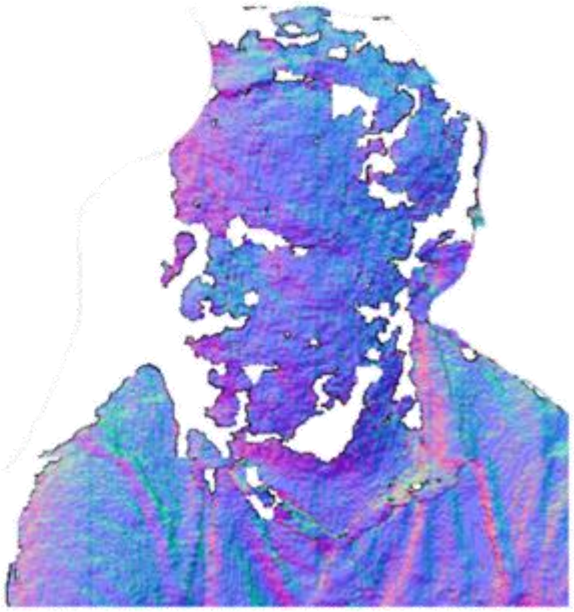
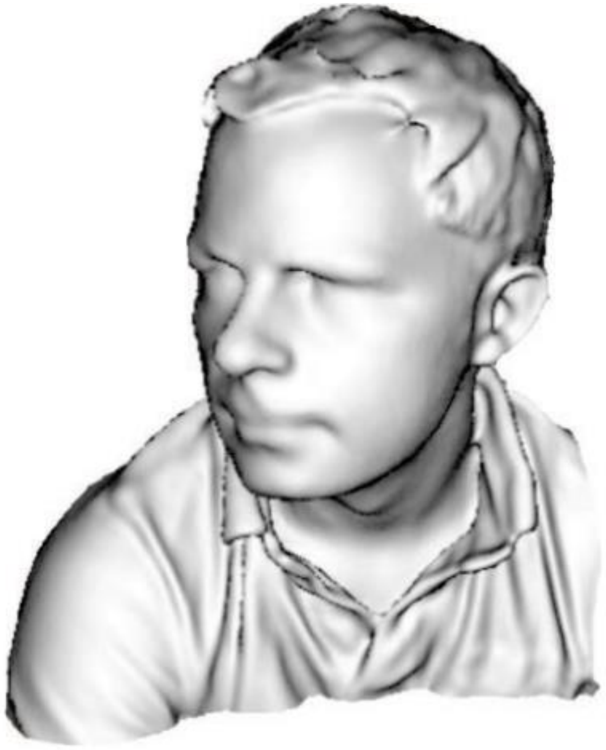


Image credit: Newcombe

KinectFusion: ICP-based Camera Tracking



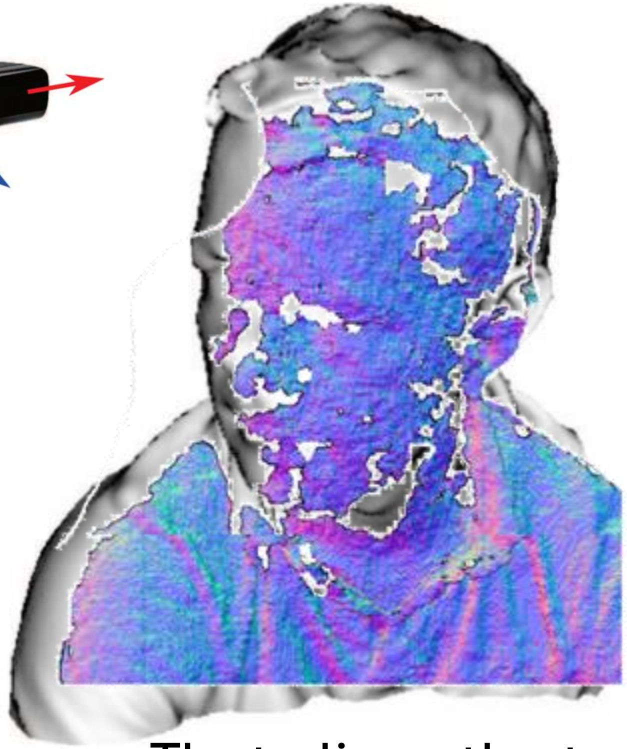
New Observation



Known 3D Model

Find the Rigid Pose

$\mathbf{T}_{w,k-1}$



That aligns the two

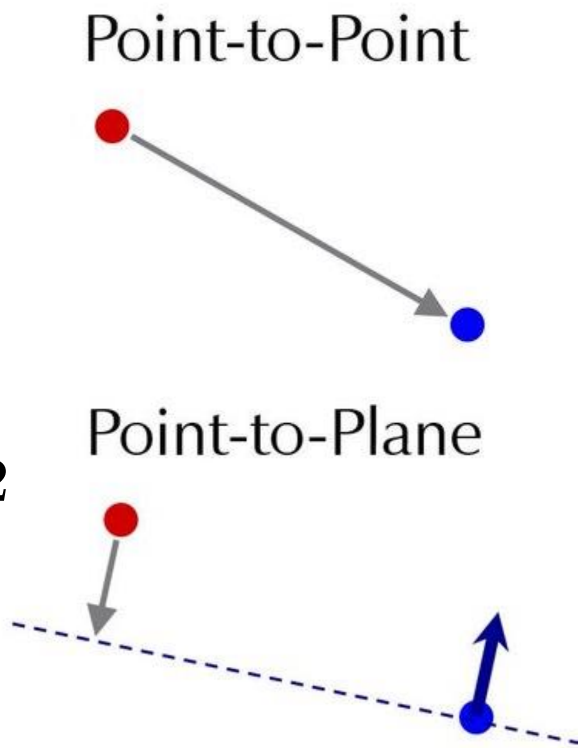
Image credit: Newcor

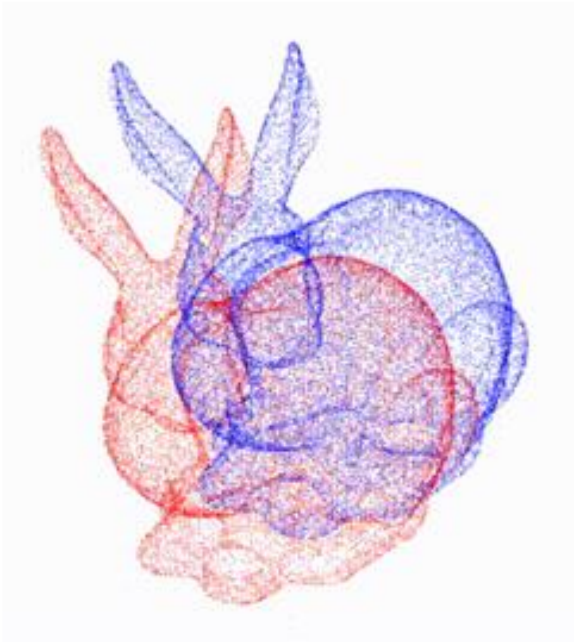
KinectFusion: ICP-based Camera Tracking

Finding the rigid transform that minimize the average distance between the two set of point cloud.

$$\min_{\mathbf{T}} \sum_{i}^{N} \|\mathbf{T}\mathbf{p}_i - \mathbf{q}_i\|_2^2$$

$$\min_{\mathbf{T}} \sum_{i}^{N} ((\mathbf{T}\mathbf{p}_i - \mathbf{q}_i) \cdot \mathbf{n}_{\mathbf{q}_i})^2$$





$\mathbf{q}_i \in \mathcal{Q}$ is the closest point to $\mathbf{T}\mathbf{p}_i$

KinectFusion: ICP-based Camera Tracking

Finding the rigid transform that minimize the average distance between the two set of point cloud.

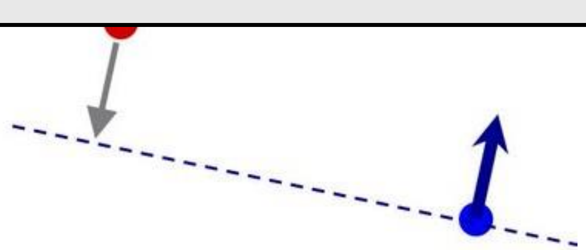
$$\min_{\mathbf{T}} \sum_{i}^{N} \|$$

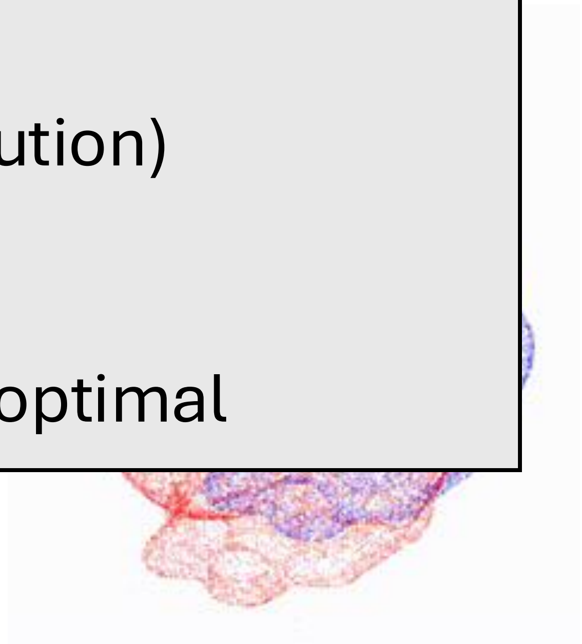
For each iteration:

- Retrieve closest point q_i for all p_i
- Update rigid transform T (closed-form solution)
- Apply T to source point cloud P

Guaranteed converge, but may not be global optimal

$$\min_{\mathbf{T}} \sum_{i}^{N} ((\mathbf{T}\mathbf{p}_i - \mathbf{q}_i)^\top \mathbf{n}_{\mathbf{q}_i})$$



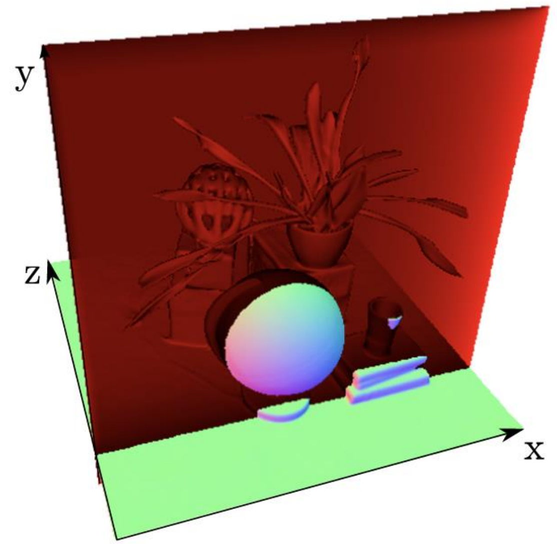


$\mathbf{q}_i \in \mathcal{Q}$ is the closest point to $\mathbf{T}\mathbf{p}_i$

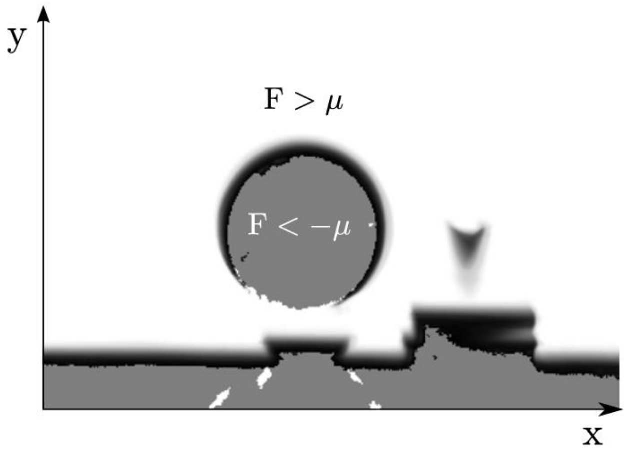
KinectFusion: TSDF

Truncated Sign Distance Function

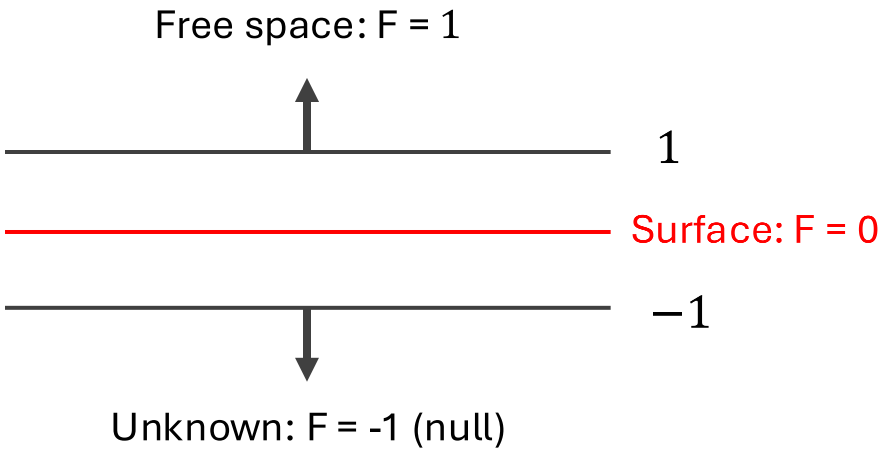
Microsoft Kinect



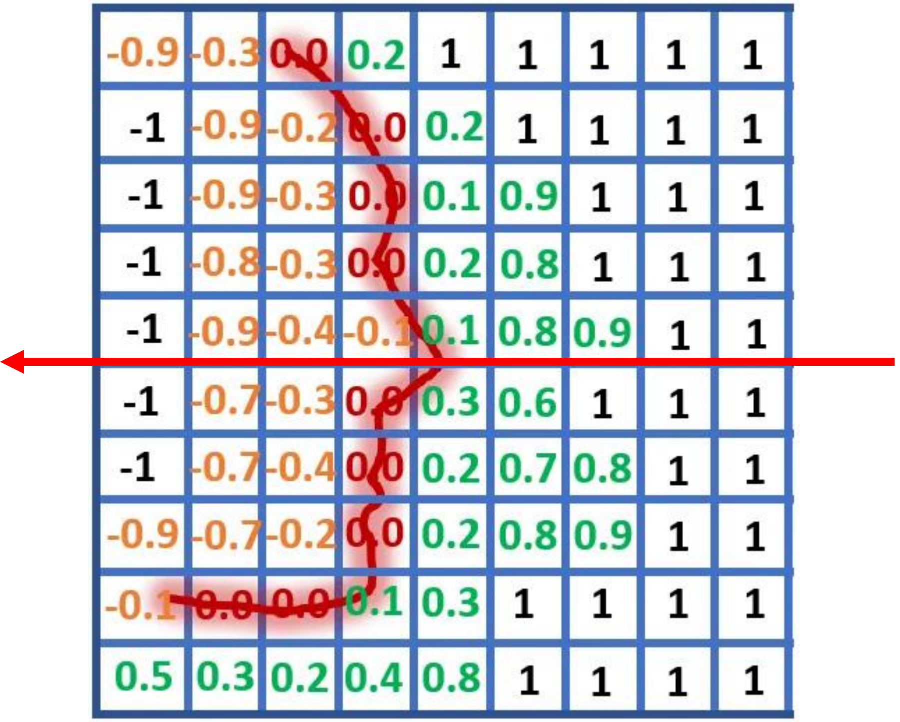
3D scene



2D slice



KinectFusion: TSDF



-0.9	-0.3	0.0	0.2	1	1	1	1	1
-1	-0.9	-0.2	0.0	0.2	1	1	1	1
-1	-0.9	-0.3	0.0	0.1	0.9	1	1	1
-1	-0.8	-0.3	0.0	0.2	0.8	1	1	1
-1	-0.9	-0.4	-0.1	0.1	0.8	0.9	1	1
-1	-0.7	-0.3	0.0	0.3	0.6	1	1	1
-1	-0.7	-0.4	0.0	0.2	0.7	0.8	1	1
-0.9	-0.7	-0.2	0.0	0.2	0.8	0.9	1	1
-0.1	0.0	0.0	0.1	0.3	1	1	1	1
0.5	0.3	0.2	0.4	0.8	1	1	1	1

2D TSDF

Measured depth — Point distance to sensor

$$\delta(p) = D(\pi(p)) - ||\,p - c||_2$$

Threshold

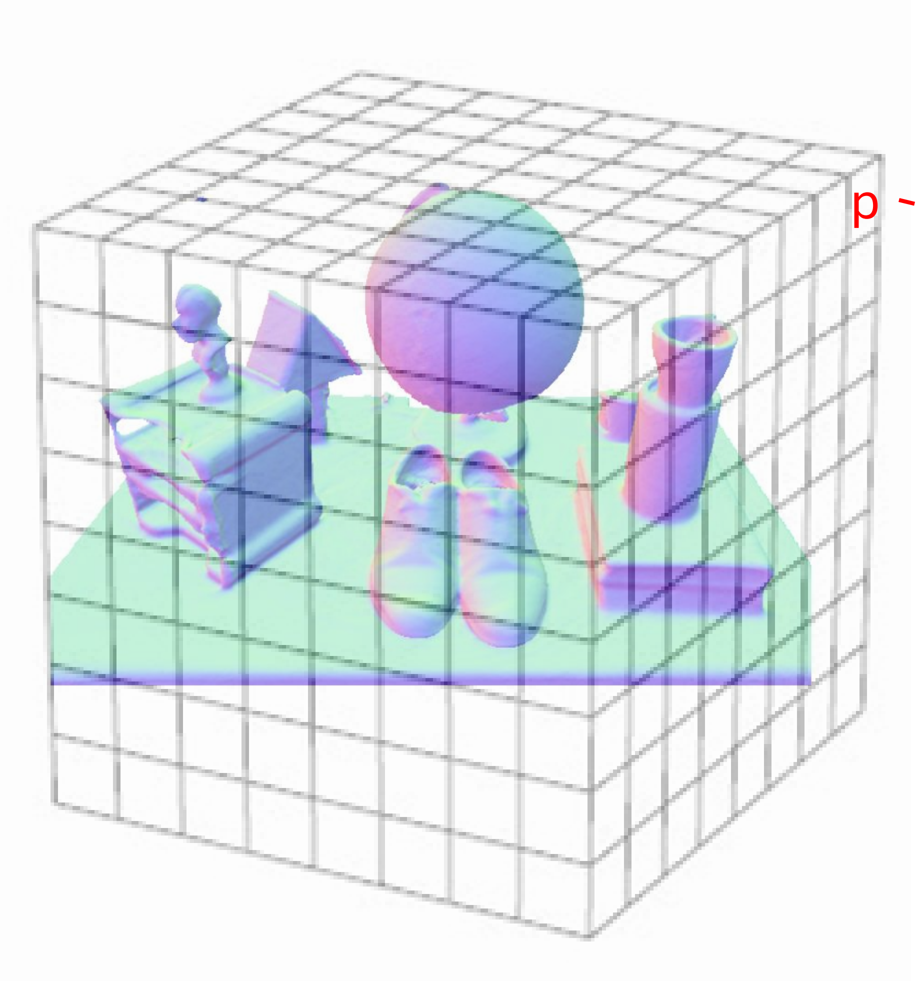
TSDF

$$F(p) = \begin{cases} \dfrac{\delta(p)}{\mu}, if\ |\delta(p)| \leq \mu \\ sign\big(\delta(p)\big), if\ |\delta(p)| > \mu \end{cases}$$

Image credit: https://itnext.io/understanding-real-time-3d-reconstruction-and-kinectfusion-33d61d1cd402

KinectFusion: TSDF

Store values in each voxel grid



$$\mathbf{S}_k(\mathbf{p}) \mapsto [\mathrm{F}_k(\mathbf{p}), \mathrm{W}_k(\mathbf{p})]$$

TSDF

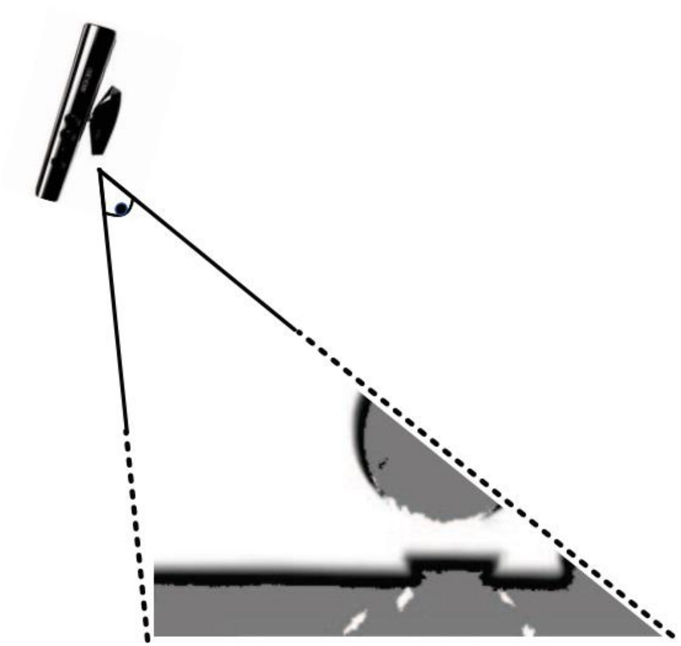
Weight

$\propto cos\theta / D(\pi(p))$

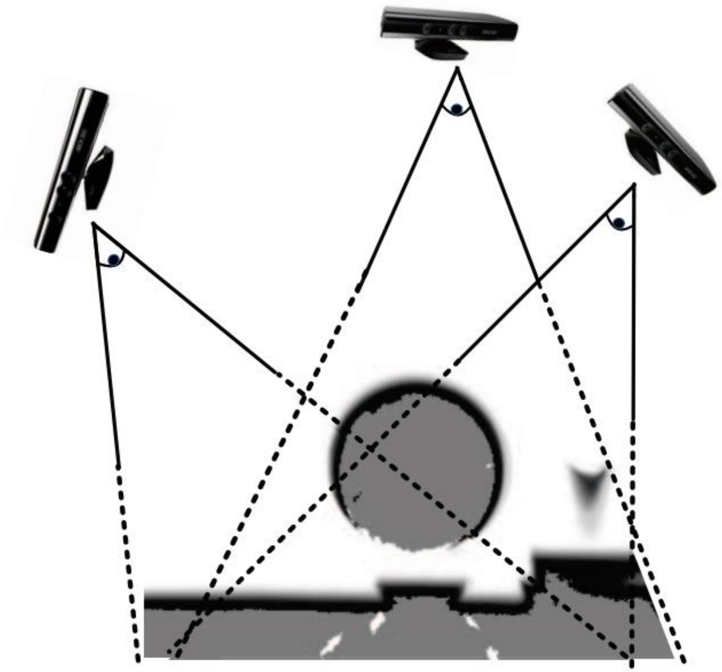
👉 Weight is high if the sensor is near and perpendicular

KinectFusion: Volume Integration

Each sensor provides a partial (noisy) observation of the signed distance function



Volume integration recursively update the map by integrating new observation.



KinectFusion: Volume Integration

Updated TSDF

Old TSDF

New Observation

$$\mathrm{F}_k(\mathbf{p}) = \frac{\mathrm{W}_{k-1}(\mathbf{p})\mathrm{F}_{k-1}(\mathbf{p}) + \mathrm{W}_{\mathrm{R}_k}(\mathbf{p})\mathrm{F}_{\mathrm{R}_k}(\mathbf{p})}{\mathrm{W}_{k-1}(\mathbf{p}) + \mathrm{W}_{\mathrm{R}_k}(\mathbf{p})}$$

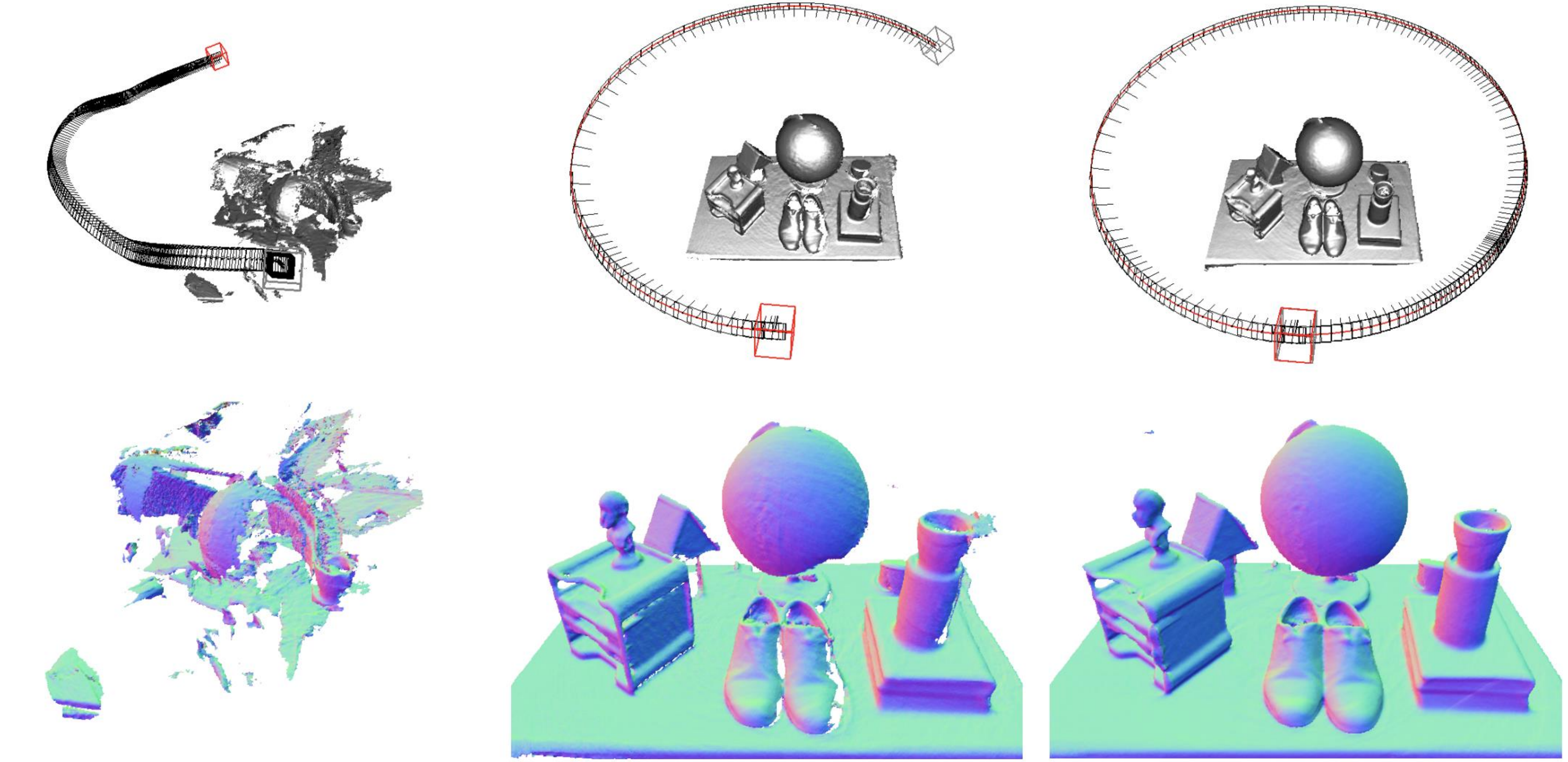
$$\mathrm{W}_k(\mathbf{p}) = \mathrm{W}_{k-1}(\mathbf{p}) + \mathrm{W}_{\mathrm{R}_k}(\mathbf{p})$$

Weight

Running weighted average

KinectFusion: Volume Integration

Frame-to-model alignment is more stable and accurate than frame-to-frame



(a) Frame to frame tracking (b) Partial loop (c) Full loop

KinectFusion: Results

The Past and the Future of KinectFusion

Archeologist: Shenlong Wang

Past and Future

- What was the status of VSLAM at the time of KinectFusion?
- What opportunities at that time enabled KinectFusion?
- What was the academic impact of KinectFusion afterwards?
- What was the industrial impact of KinectFusion?

VSLAM in 2011: Current State

Real-time camera tracking and mapping (e.g. PTAM):

- Real-time and robustness in camera tracking is possible
- Maps are represented as a set of sparse feature points
- No real-time dense reconstruction

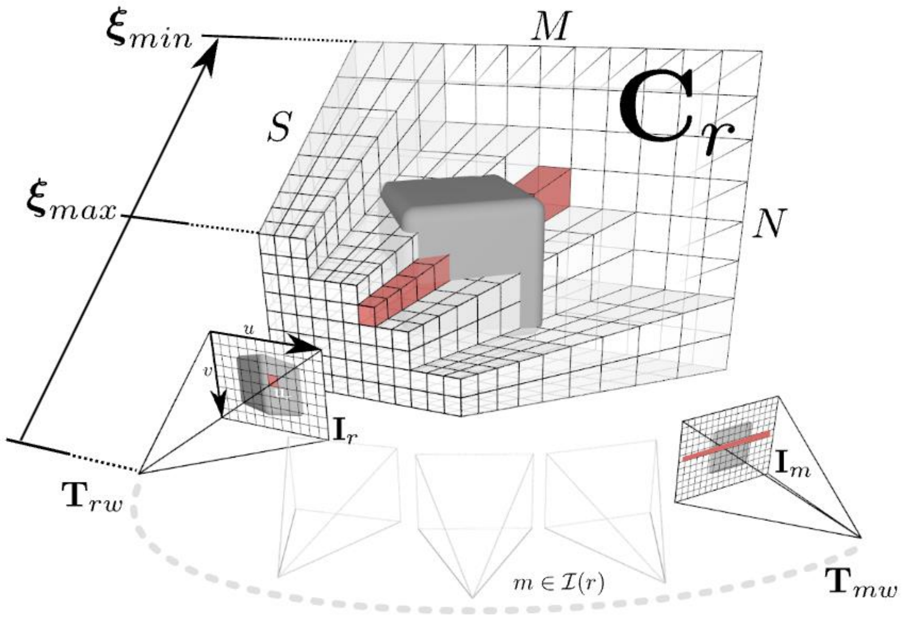
Why dense map is appealing?

Klein and Murray, Parallel Tracking and Mapping for Small AR Workspaces ISMAR 2007

VSLAM in 2011: Current State

Dense Tracking and Mapping (DTAM)!

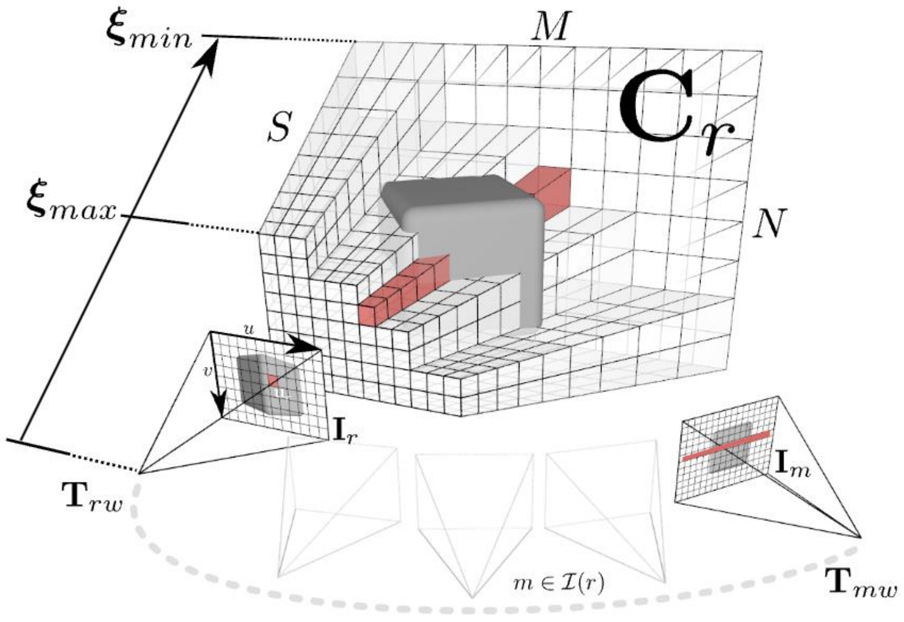
- Maps --> dense cost volume at keyframes.



DTAM: Dense Tracking and Mapping in Real-Time

Richard A. Newcombe, Steven J. Lovegrove and Andrew J. Davison
Department of Computing, Imperial College London, UK
[rnewcomb,sl203,ajd]@doc.ic.ac.uk

VSLAM in 2011: Current State



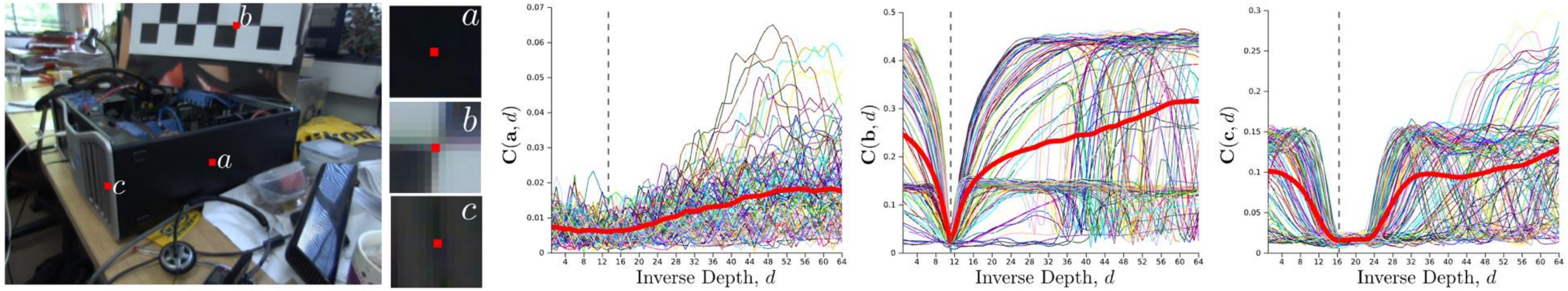
Dense Tracking and Mapping (DTAM)!

- Maps --> dense cost volume at keyframes.
- Cost: sum of photometric error error:

$$\mathbf{C}_r(\mathbf{u}, d) = \frac{1}{|\mathcal{I}(r)|} \sum_{m \in \mathcal{I}(r)} \|\rho_r(\mathbf{I}_m, \mathbf{u}, d)\|_1$$

$$\rho_r(\mathbf{I}_m, \mathbf{u}, d) = \mathbf{I}_r(\mathbf{u}) - \mathbf{I}_m\left(\pi\left(\mathrm{K}\mathrm{T}_{mr}\pi^{-1}(\mathbf{u}, d)\right)\right).$$

Intuition: 1. warping image using true depth value yield smaller photometric error (brightness consistency)
2. averaging overall multiple images makes the cost to be robust.



DTAM:
Dense Tracking and Mapping in Real-Time

VSLAM in 2011: Current State

- Real-time Sparse SLAM method doesn’t provide dense reconstruction of the scene.
- Dense SLAM starts emerging and provides dense 3D maps, do not provide truly real-time dense recon, and are limited under certain illumination.
- Desiderata: real-time, dense, robust

Klein and Murray, Parallel Tracking and Mapping for Small AR Workspaces ISMAR 2007

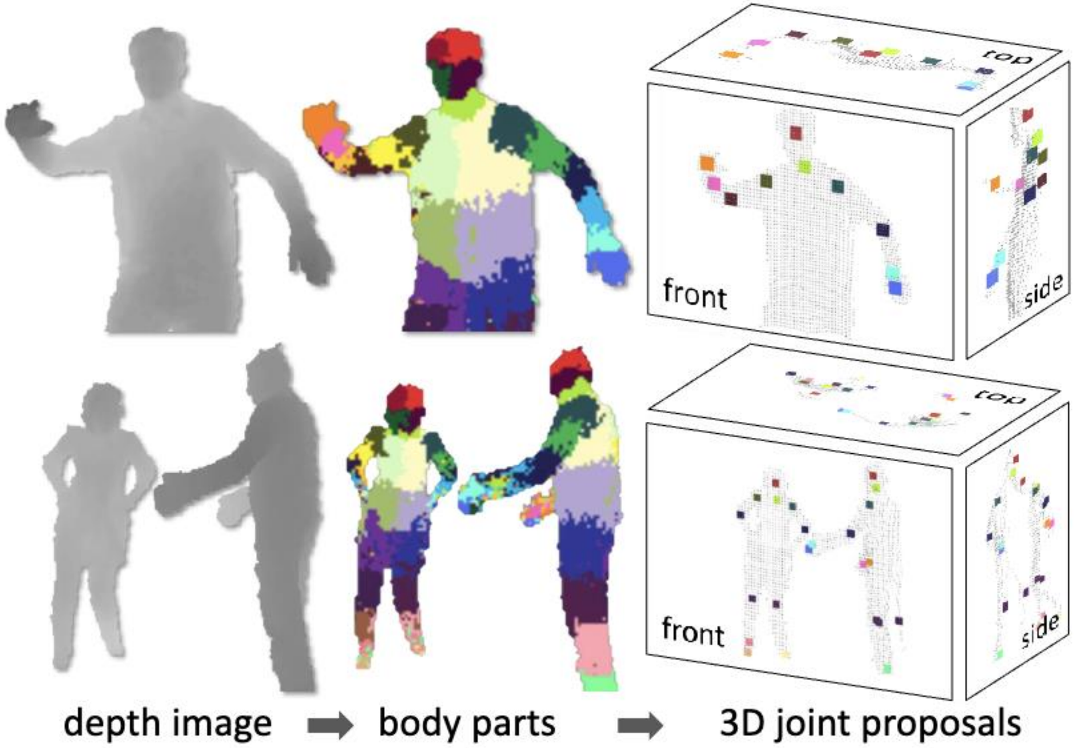
Opportunities in 2011

- Sensor: Kinect is out!

Image credit: BBC

Real-Time Human Pose Recognition in Parts from Single Depth Images

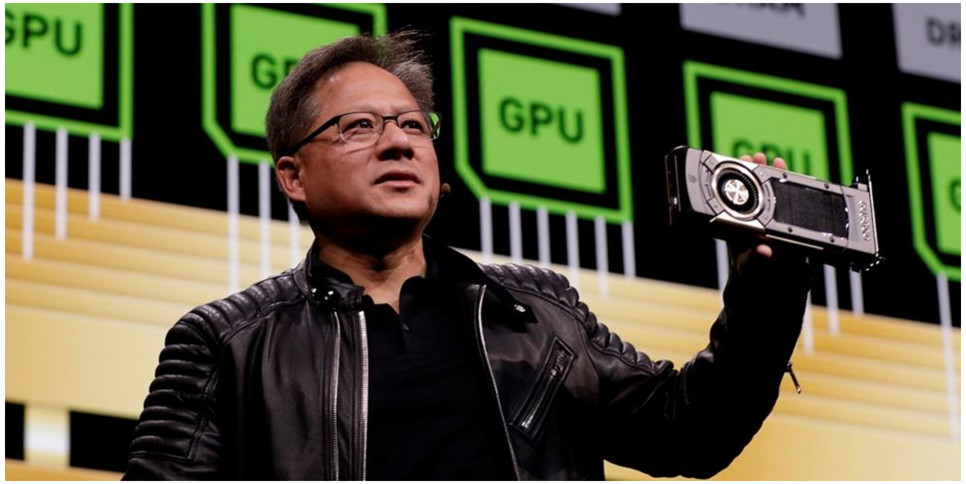
Jamie Shotton   Andrew Fitzgibbon   Mat Cook   Toby Sharp   Mark Finocchio
Richard Moore   Alex Kipman   Andrew Blake
Microsoft Research Cambridge & Xbox Incubation



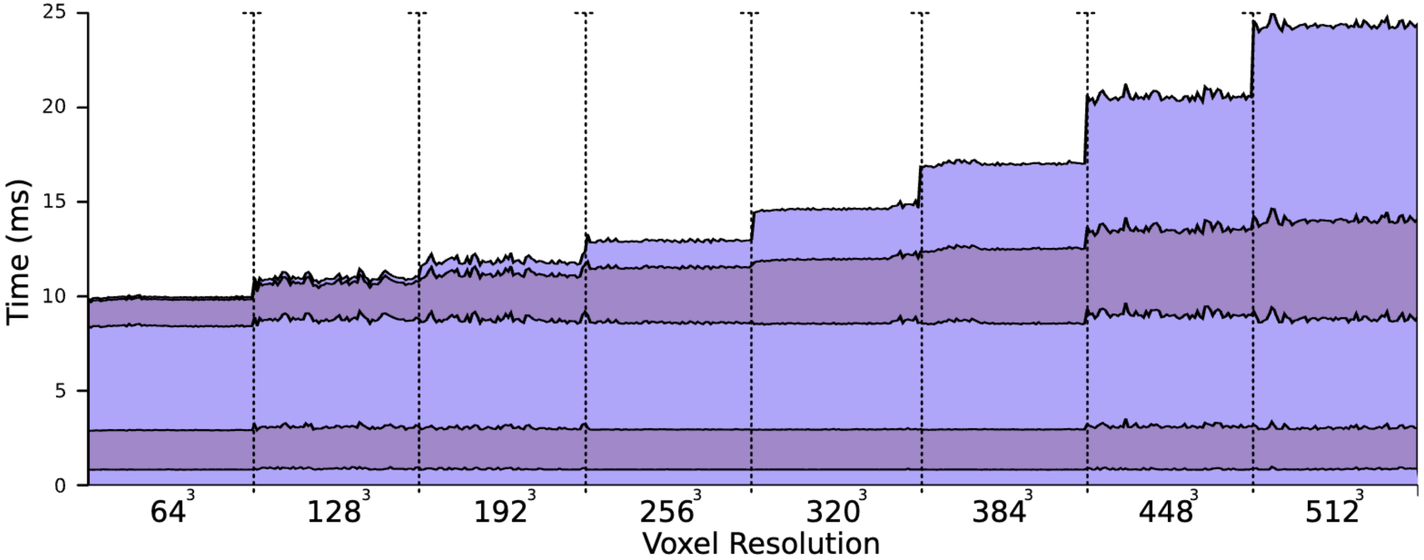
Shotton et al. *Real-time human pose recognition in parts from single depth images*, CVPR 2011 Best Paper.

Opportunities in 2011

- GPGPUs and CUDA library
- TSDF fusion and normal estimation are highly parallel computation
- GPU helps a ton



Nvidia



Real-Time Human Pose Recognition in Parts from Single Depth Images

CVPR 2011

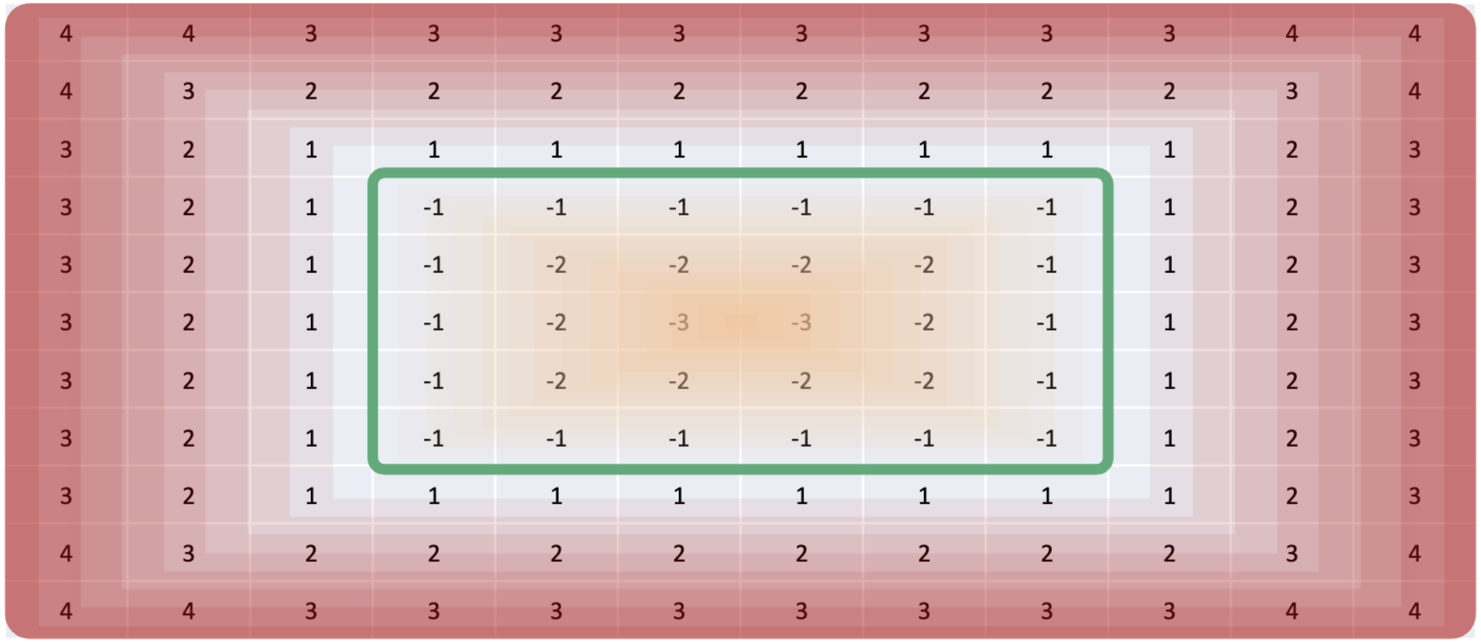
Jamie Shotton, Andrew Fitzgibbon, Mat Cook, Toby Sharp,
Mark Finocchio, Richard Moore, Alex Kipman, Andrew Blake

Microsoft Research

Xbox Incubation

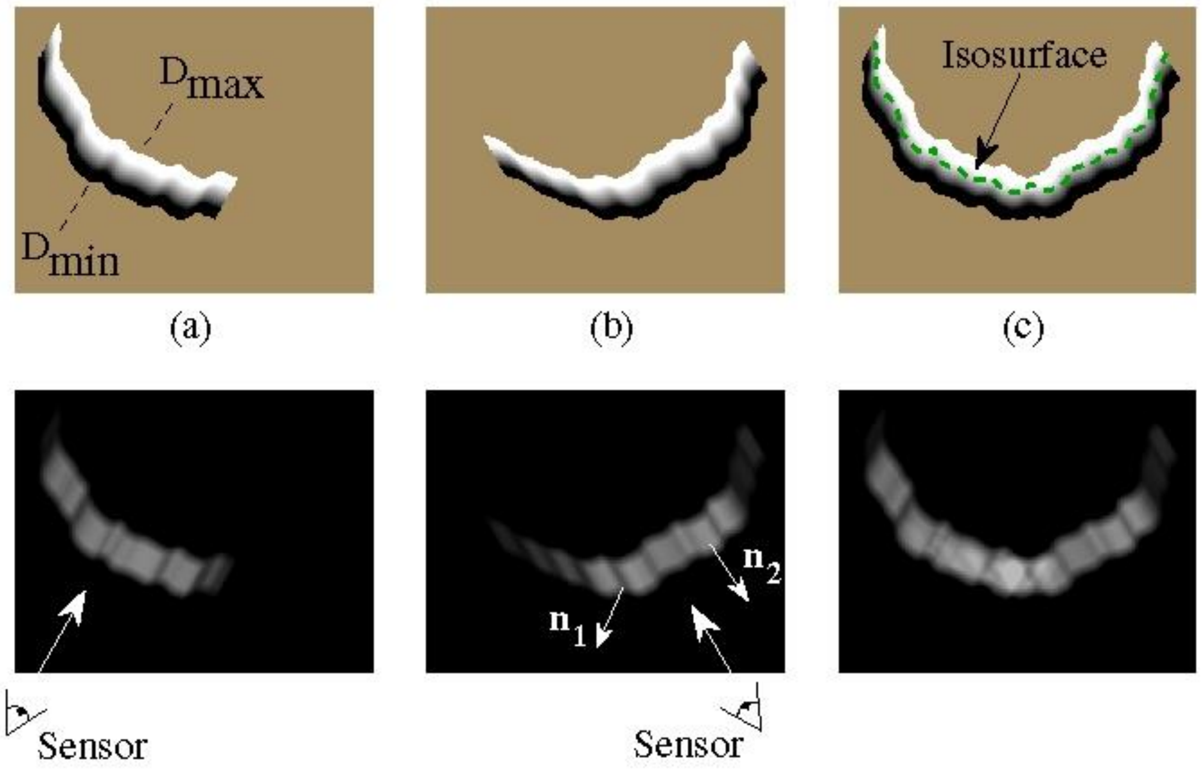
Opportunity: Volume Integration

- Signed distance function to represent 3D geometry
- Value – distance of a point to its closet surface
- Positive – outside surface, Negative – inside surface, Zero -- Isosurface



Opportunity: Volume Integration

- Good news: partial SDF from multiple scans can be integrated



(a) (b) (c) (d) (e) (f)

$$D_{i+1}(\mathbf{x}) = \frac{W_i(\mathbf{x})D_i(\mathbf{x}) + w_{i+1}(\mathbf{x})d_{i+1}(\mathbf{x})}{W_i(\mathbf{x}) + w_{i+1}(\mathbf{x})}$$

$$W_{i+1}(\mathbf{x}) = W_i(\mathbf{x}) + w_{i+1}(\mathbf{x})$$

Curless and Levoy, A Volumetric Method for Building Complex Models from Range Images, SIGGRAPH 96

KinectFusion → what happened next?

Scale up to very large scenes and faster

- Key idea: use hashing to store only occupied voxels

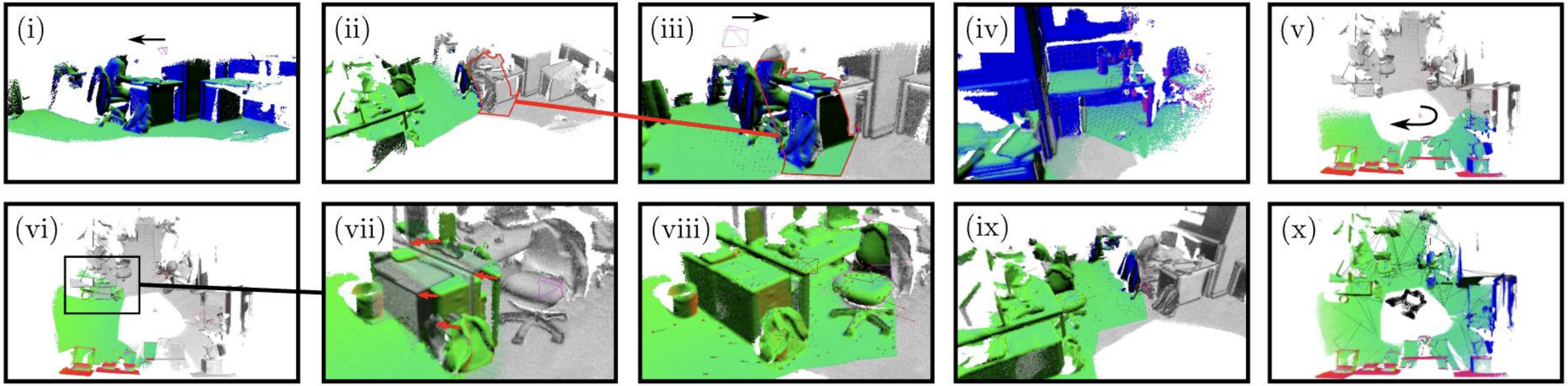
Real-time 3D Reconstruction at Scale using Voxel Hashing

Matthias Nießner[1,3] Michael Zollhöfer[1] Shahram Izadi[2] Marc Stamminger[1]
[1]University of Erlangen-Nuremberg [2]Microsoft Research Cambridge [3]Stanford University

KinectFusion → what happened next?

Solving drifting and global consistency

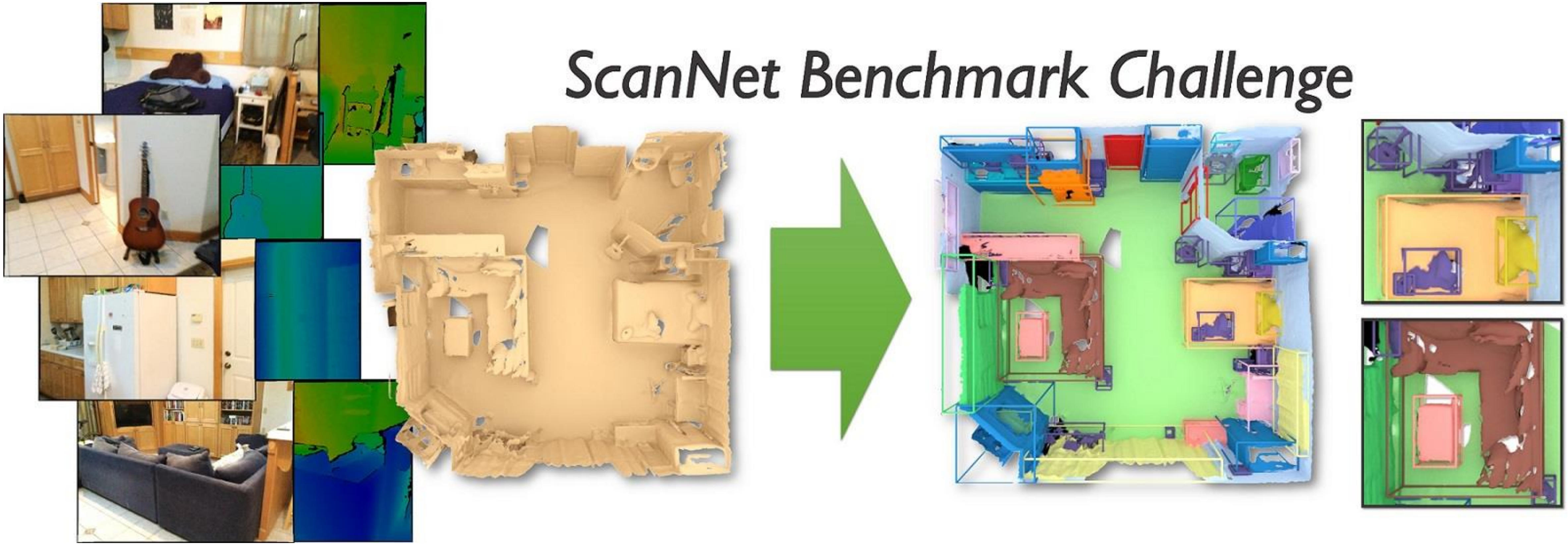
- Key idea: represent the 3D with explicit geometry primitives so we can directly "deform" locally to align globally



Whelan et al. ElasticFusion: Real-Time Dense SLAM and Light Source Estimation, IJRR 16

KinectFusion → what happened next?

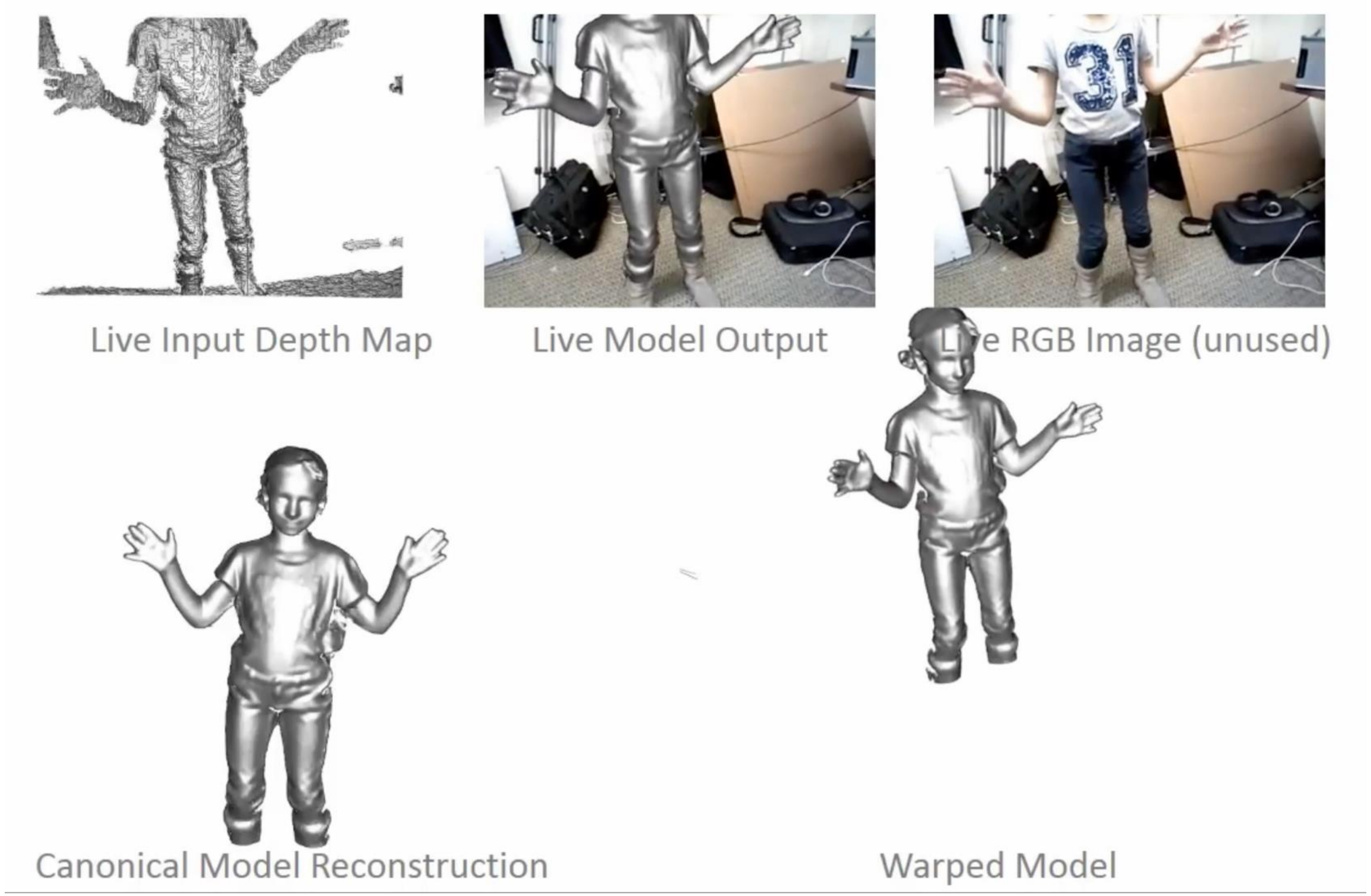
- Data engine for many tasks: e.g. 3D scene understanding



Dai et al. ScanNet: Richly-annotated 3D Reconstructions of Indoor Scenes, CVPR 2017

KinectFusion → what happened next?

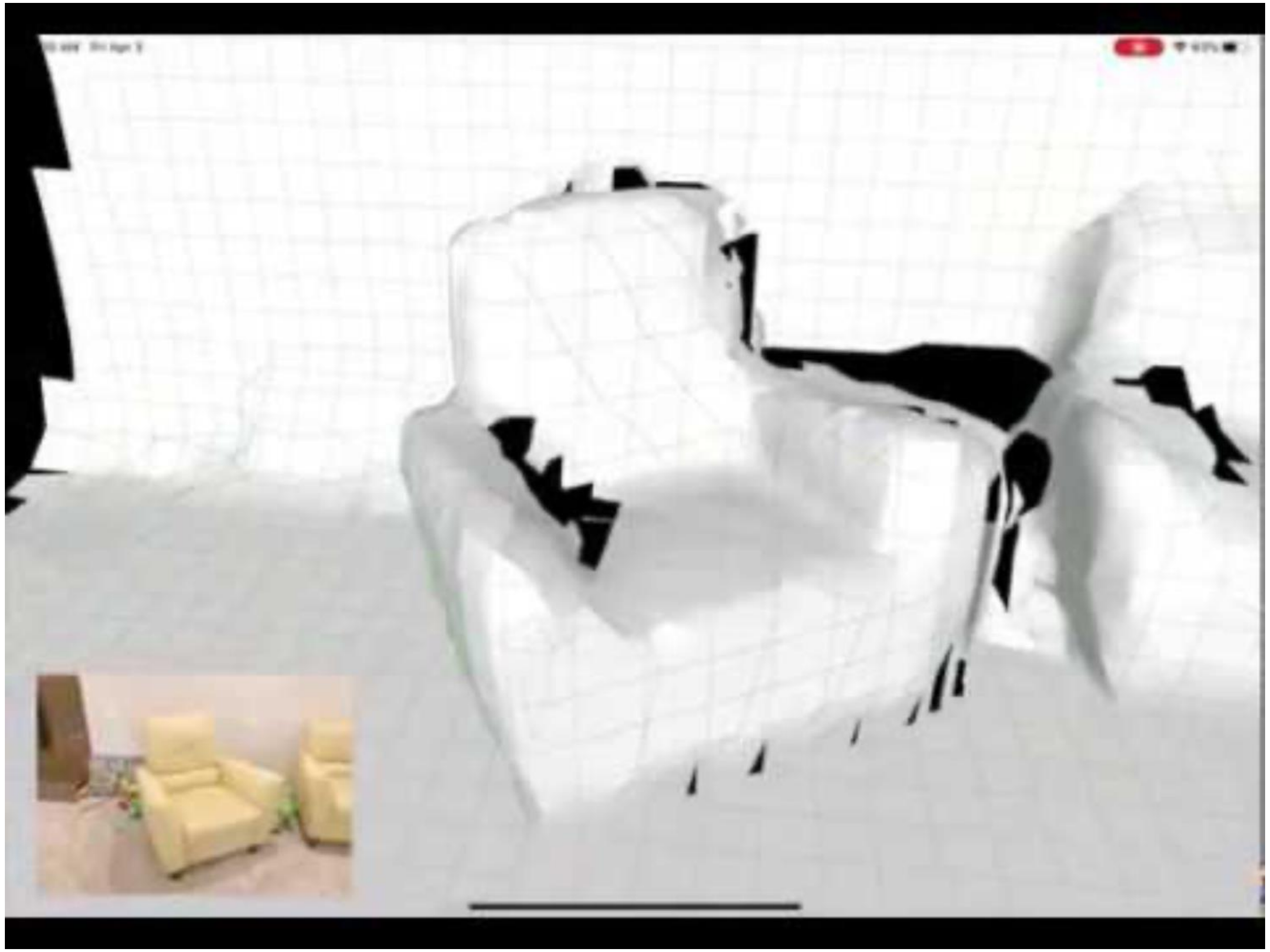
- Extended to dynamic scenes



Newcombe, et al. DynamicFusion: Reconstruction and Tracking of Non-rigid Scenes in Real-Time, CVPR 2015

KinectFusion → what happened next?

- Real-time 3D reconstruction product in iPhone



ARKit for 3D Reconstruction

VSLAM → what happened next?

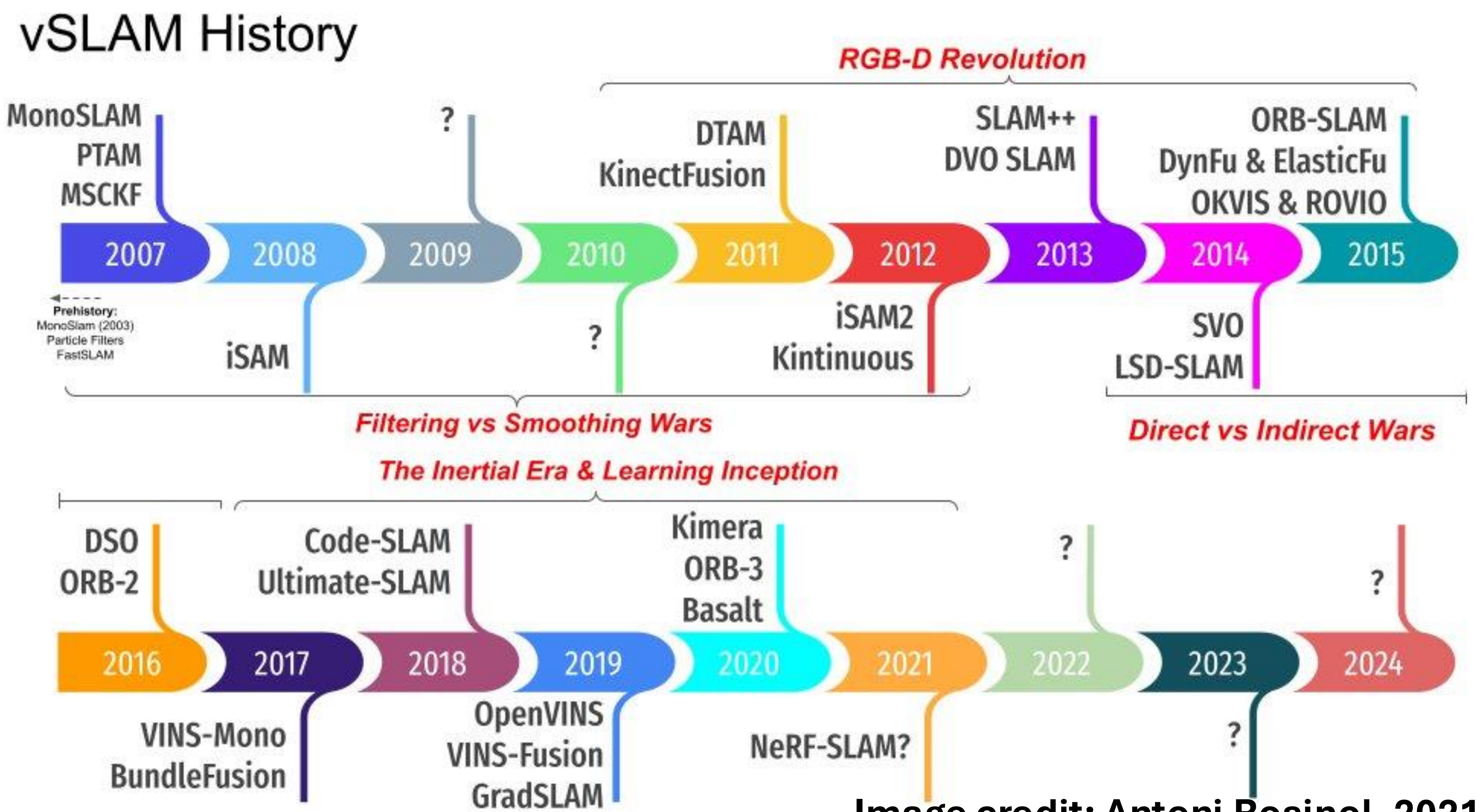


Image credit: Antoni Rosinol, 2021

Andrew Davison:

MonoSLAM,
KinectFusion,
DTAM,
SLAM++,
ElasticFusion,
SemanticFusion,
CodeSLAM,
iMap,
Gaussian splatting slam

Good survey: Cadena, Past, present, and future of simultaneous localization and mapping: Toward the robust-perception age, TRO, 2016

Private Investigator: Richard Newcombe

Shenlong Wang

Richard Newcombe

- PhD: Imperial College at London
- Advisor: Andrew J. Davison

VP, Research Science
Meta · Full-time
Aug 2022 - Present · 2 yrs 2 mos
United States

Facebook
4 yrs 9 mos

Director of Research Science
2018 - Sep 2022 · 4 yrs 9 mos

Director of Research Science
Jan 2018 - Aug 2022 · 4 yrs 8 mos

Oculus VR
4 yrs 8 mos

Director, Research Science
Jan 2017 - Aug 2019 · 2 yrs 8 mos

Manager, Research Science
2015 - 2017 · 2 yrs
United States

Co Founder
Surreal Vision Ltd
Oct 2014 - Sep 2015 · 1 yr

Researcher
University of Washington
Dec 2012 - Sep 2015 · 2 yrs 10 mos

Research student
Imperial College London
Apr 2008 - Nov 2012 · 4 yrs 8 mos

Working in the cognitive robotics and robot vision groups.

Research intern
Microsoft Research, Cambridge
Sep 2010 - Dec 2010 · 4 mos

Richard Newcombe

- PhD: Imperial College at London
- Advisor: Andrew J. Davison

Co Founder
Surreal Vision Ltd
Oct 2014 - Sep 2015 · 1 yr

Acquired by Oculus

Researcher
University of Washington
Dec 2012 - Sep 2015 · 2 yrs 10 mos

Postdoc: Dynamic Fusion
(we will cover later this fall)

Research student
Imperial College London
Apr 2008 - Nov 2012 · 4 yrs 8 mos

Working in the cognitive robotics and robot vision groups.

PhD: DTAM
Super-well written thesis

Research intern
Microsoft Research, Cambridge
Sep 2010 - Dec 2010 · 4 mos

Intern project: KinectFusion

VP, Research Science
Meta · Full-time
Aug 2022 - Present · 2 yrs 2 mos
United States

Replica, Ego4D

Facebook
4 yrs 9 mos

DeepSDF
(we will cover later this fall)

Director of Research Science
2018 - Sep 2022 · 4 yrs 9 mos

Director of Research Science
Jan 2018 - Aug 2022 · 4 yrs 8 mos

Oculus VR
4 yrs 8 mos

Acquired by Facebook

Director, Research Science
Jan 2017 - Aug 2019 · 2 yrs 8 mos

Manager, Research Science
2015 - 2017 · 2 yrs
United States

Richard Newcombe

Lesson: #1: Write an excellent PhD thesis that people can learn from it

Excellent literature review for V-SLAM

Well-written tutorial on Dense SLAM

3D Vision and Optimization 101

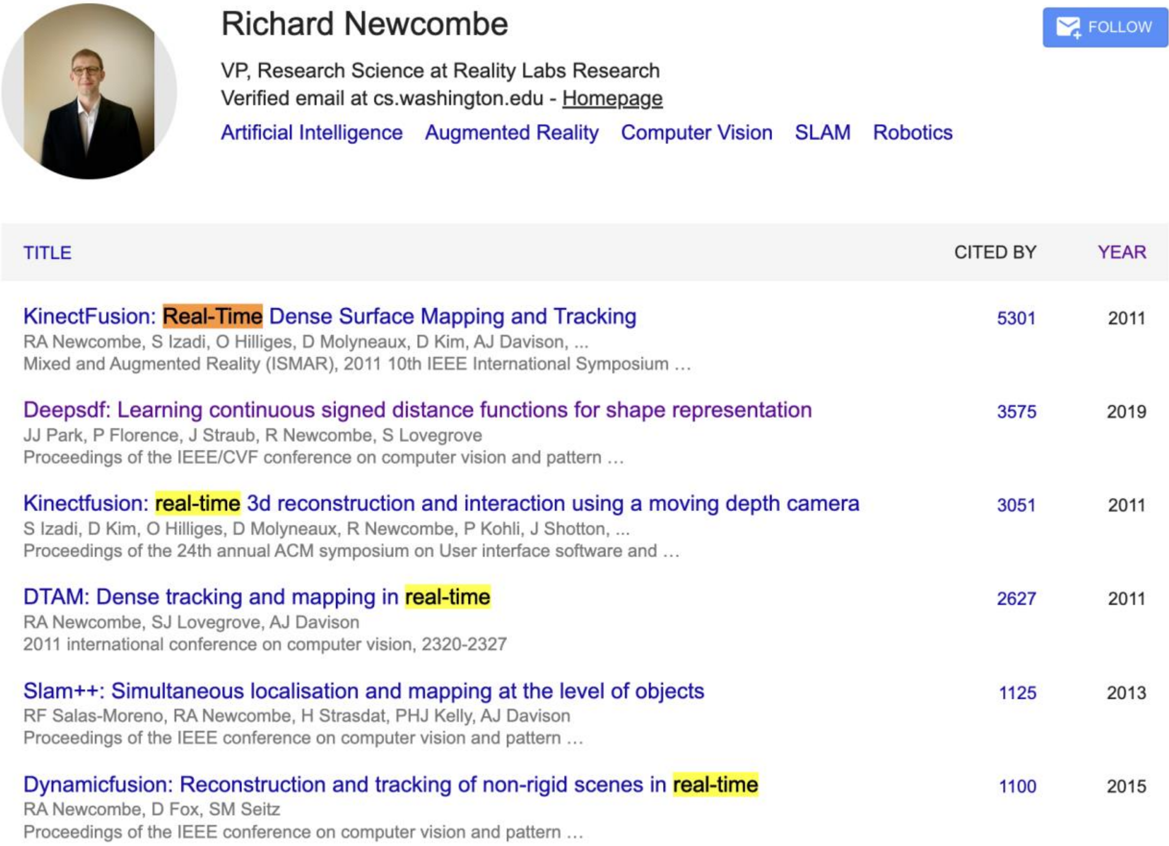
Primal dual for depth denoising

Dense Visual SLAM Richard Newcombe 2012 PhD thesis

Richard Newcombe

Lesson: #2: the joy of live demos that work

- Live demo: easier for people to run, reproduce and appreciate
- Real-time: easier for people to turn into actual products
- System-level optimization: a great testament to your coding skills


Richard Newcombe
FOLLOW
VP, Research Science at Reality Labs Research
Verified email at cs.washington.edu - Homepage
Artificial Intelligence Augmented Reality Computer Vision SLAM Robotics

TITLE	CITED BY	YEAR
KinectFusion: Real-Time Dense Surface Mapping and Tracking RA Newcombe, S Izadi, O Hilliges, D Molyneaux, D Kim, AJ Davison, ... Mixed and Augmented Reality (ISMAR), 2011 10th IEEE International Symposium ...	5301	2011
Deepsdf: Learning continuous signed distance functions for shape representation JJ Park, P Florence, J Straub, R Newcombe, S Lovegrove Proceedings of the IEEE/CVF conference on computer vision and pattern ...	3575	2019
Kinectfusion: real-time 3d reconstruction and interaction using a moving depth camera S Izadi, D Kim, O Hilliges, D Molyneaux, R Newcombe, P Kohli, J Shotton, ... Proceedings of the 24th annual ACM symposium on User interface software and ...	3051	2011
DTAM: Dense tracking and mapping in real-time RA Newcombe, SJ Lovegrove, AJ Davison 2011 international conference on computer vision, 2320-2327	2627	2011
Slam++: Simultaneous localisation and mapping at the level of objects RF Salas-Moreno, RA Newcombe, H Strasdat, PHJ Kelly, AJ Davison Proceedings of the IEEE conference on computer vision and pattern ...	1125	2013
Dynamicfusion: Reconstruction and tracking of non-rigid scenes in real-time RA Newcombe, D Fox, SM Seitz Proceedings of the IEEE conference on computer vision and pattern ...	1100	2015

Dense Visual SLAM Richard Newcombe 2012 PhD thesis

KinectBot

KinectFusion Product Idea

Albert Zhai

KinectBot: your helpful household robot

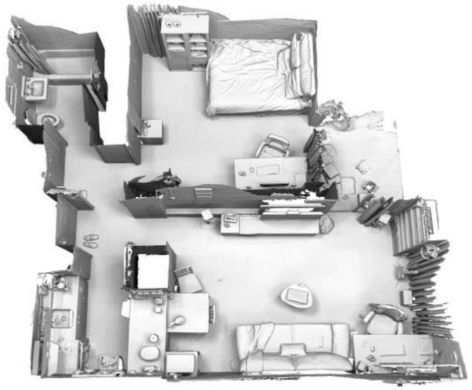
- A mobile robot with a Kinect camera and a simple arm
- Keeps track of objects around the home and performs general manipulation tasks
  - “KinectBot, bring me my keys!”
- 3D mapping powered by KinectFusion

Image credits: Microsoft, Hello Robot

KinectBot: how it works

Mapping



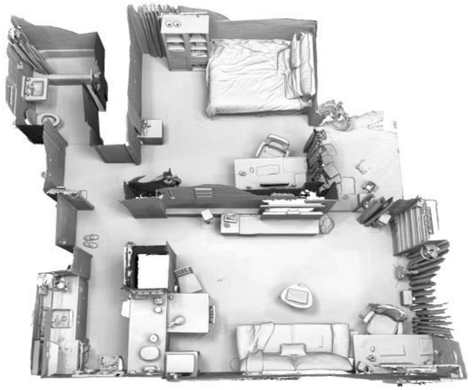


- KinectBot will tour the house every few hours
- Each time, build a 3D map using KinectFusion

Image credits: RentCafe, Qianyi Zhou, ScanNet++, KinectFusion

KinectBot: how it works

Mapping



- KinectBot will tour the house every few hours
- Each time, build a 3D map using KinectFusion

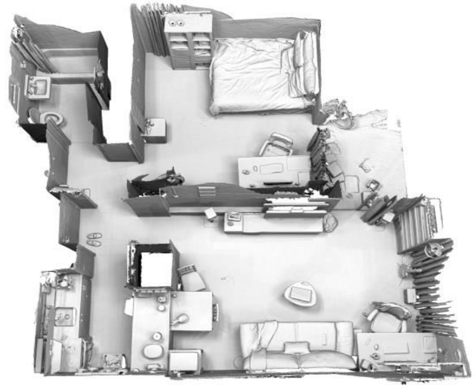
Identifying Objects

- Use KinectFusion's dynamic object segmentation abilities to find objects that move
- Gradually build up a database of all objects and their last locations

Image credits: RentCafe, Qianyi Zhou, ScanNet++, KinectFusion

KinectBot: how it works

Mapping



- KinectBot will tour the house every few hours
- Each time, build a 3D map using KinectFusion

Identifying Objects

- Use KinectFusion's dynamic object segmentation abilities to find objects that move
- Gradually build up a database of all objects and their last locations

Task Execution

"Bring me my headphones"

Planning Models

- Use off-the-shelf models and object database to perform task
  - VLM for retrieval and text understanding
  - Grasping model
  - Motion planner

Image credits: RentCafe, Qianyi Zhou, ScanNet++, KinectFusion

KinectBot: highlights and limitations

Highlights

- Real-time capability of KinectFusion allows for dynamic mapping and segmentation even when people are moving around
- Kinect system is relatively cheap

Limitations

- Relies on objects being moved to be detected
- Tours may take a while; does not update in between tours

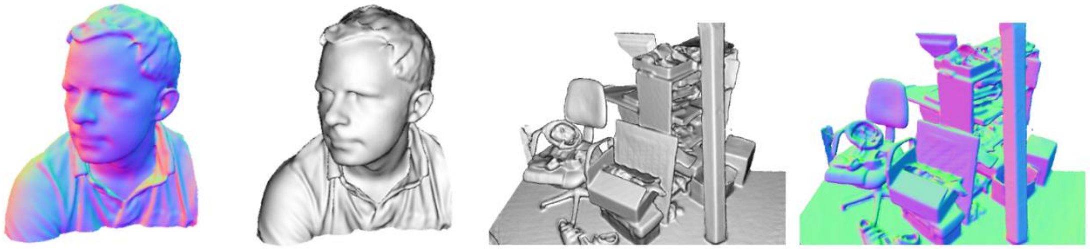


Image credits: KinectFusion, UIUC STORIED

Hacker: 2D KinectFusion

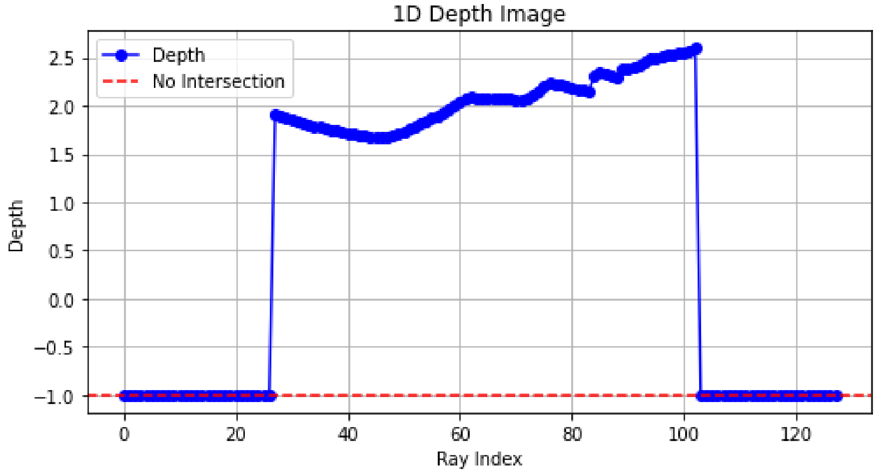
Shenlong Wang

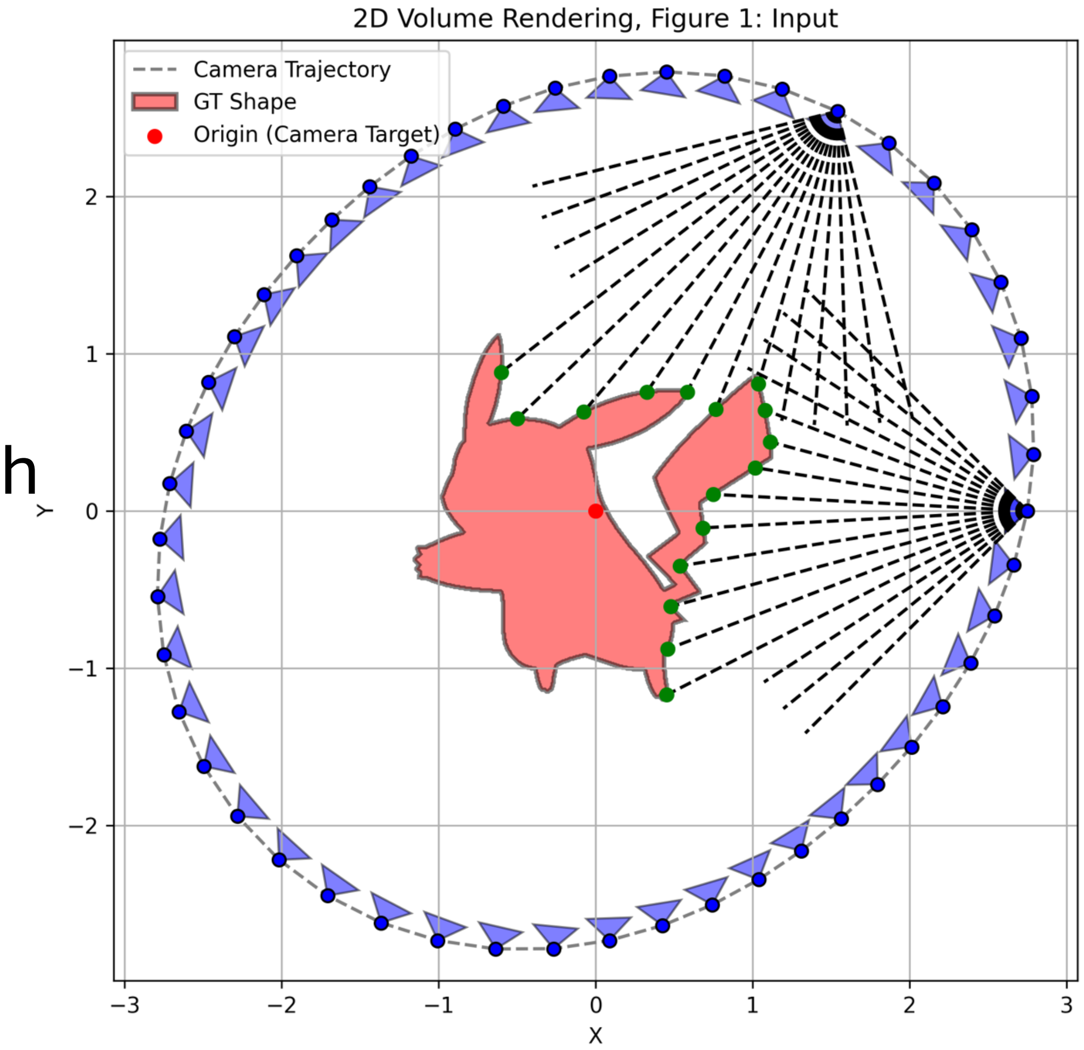
Code will be available on our shared Github repo (Zhi-Hao will share later today)

Kinect Fusion in a 2D World

- Assuming we live in a 2D world, everyone perceives the 2D world through a 1D perspective imaging.

- One day, 2D computer scientists invented a '1.5D' camera, where each pixel captures the depth of the ray.

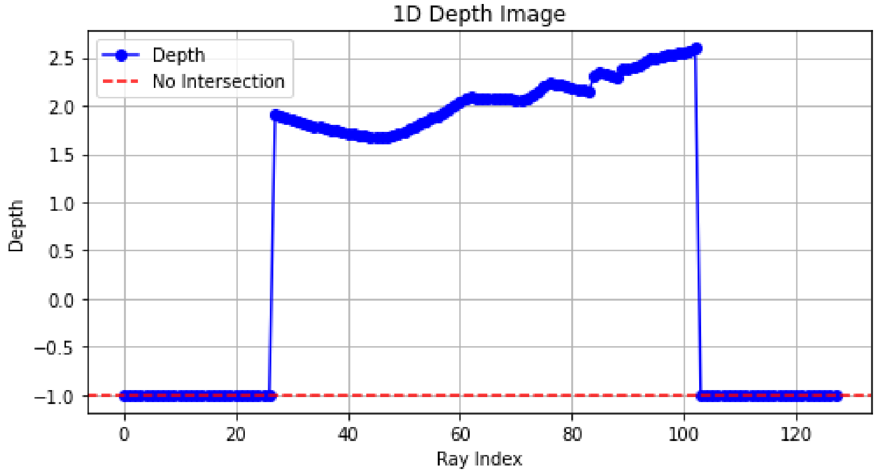


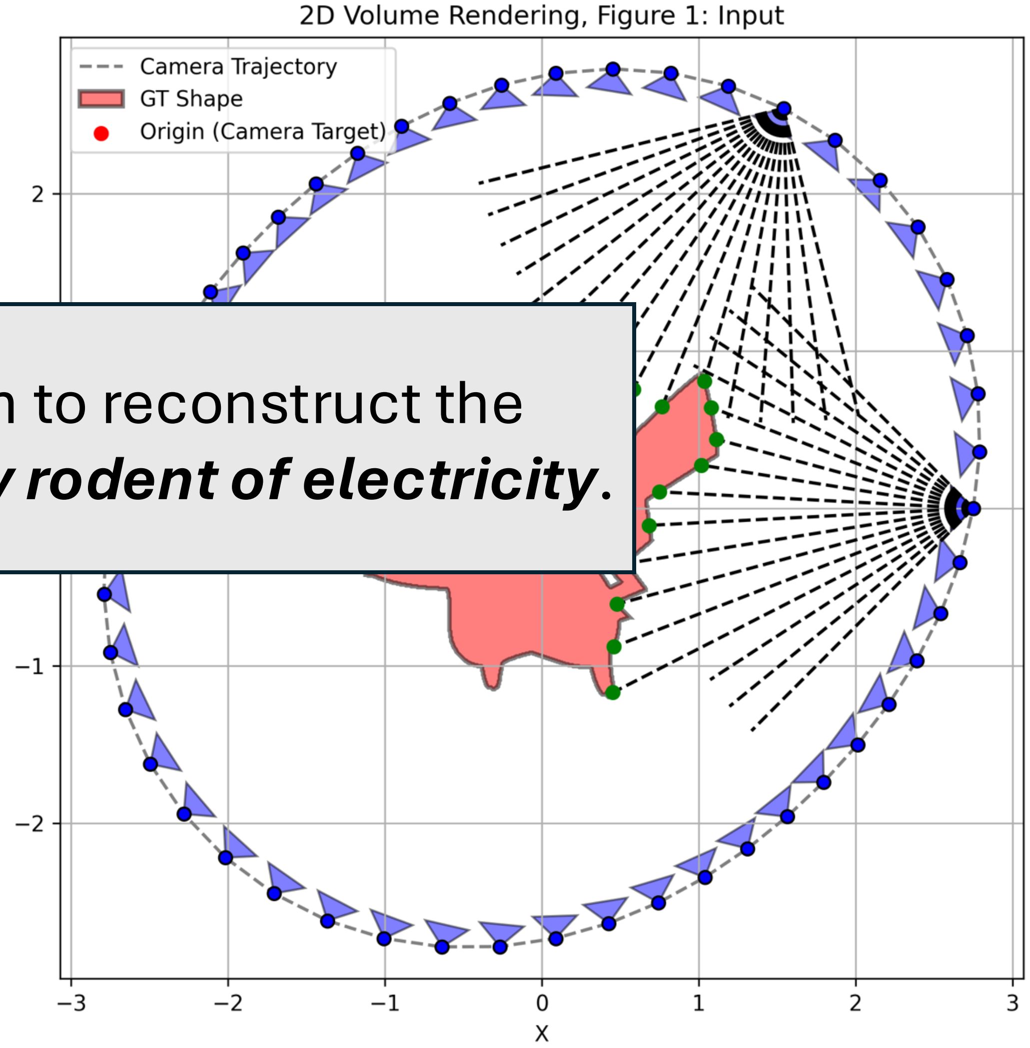


Kinect Fusion in a 2D World

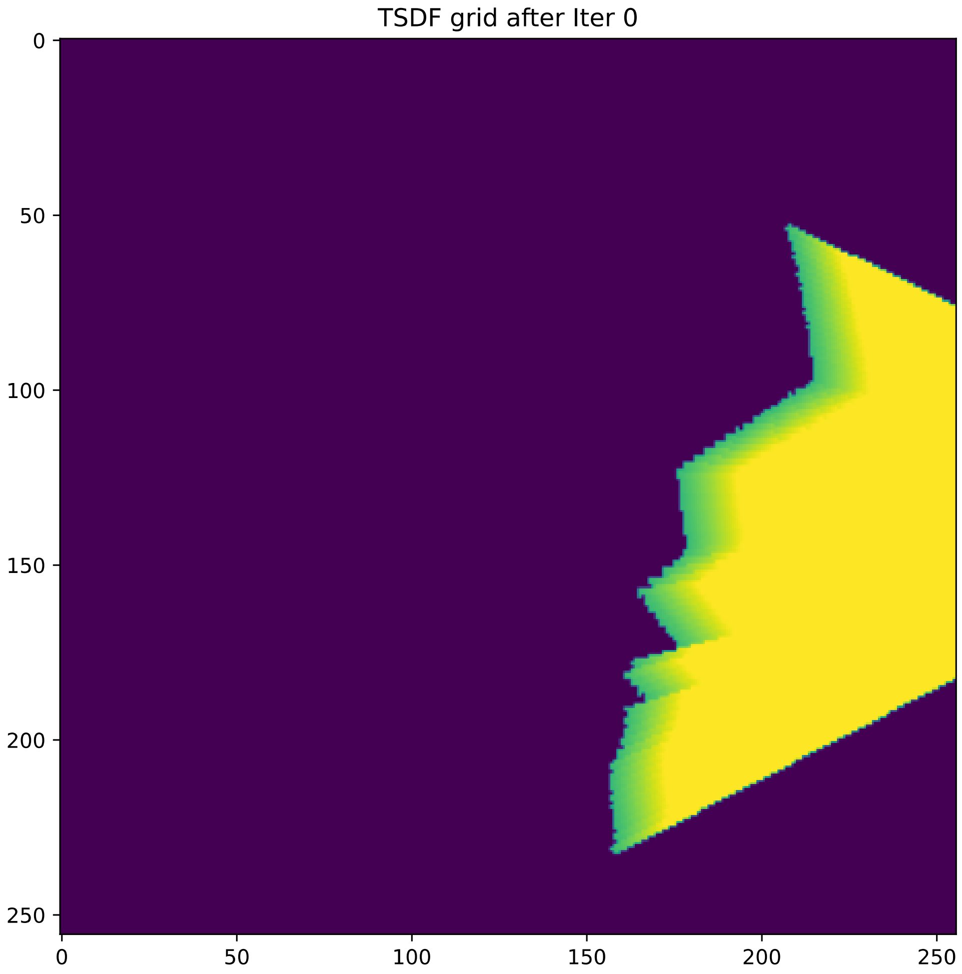
- Assuming we live in a 2D world, everyone perceives the 2D world through a 1D perspective imaging.
- One day, … invented … pixel cap…

They want to use this invention to reconstruct the shape of their giant idol – the ***holy rodent of electricity***.

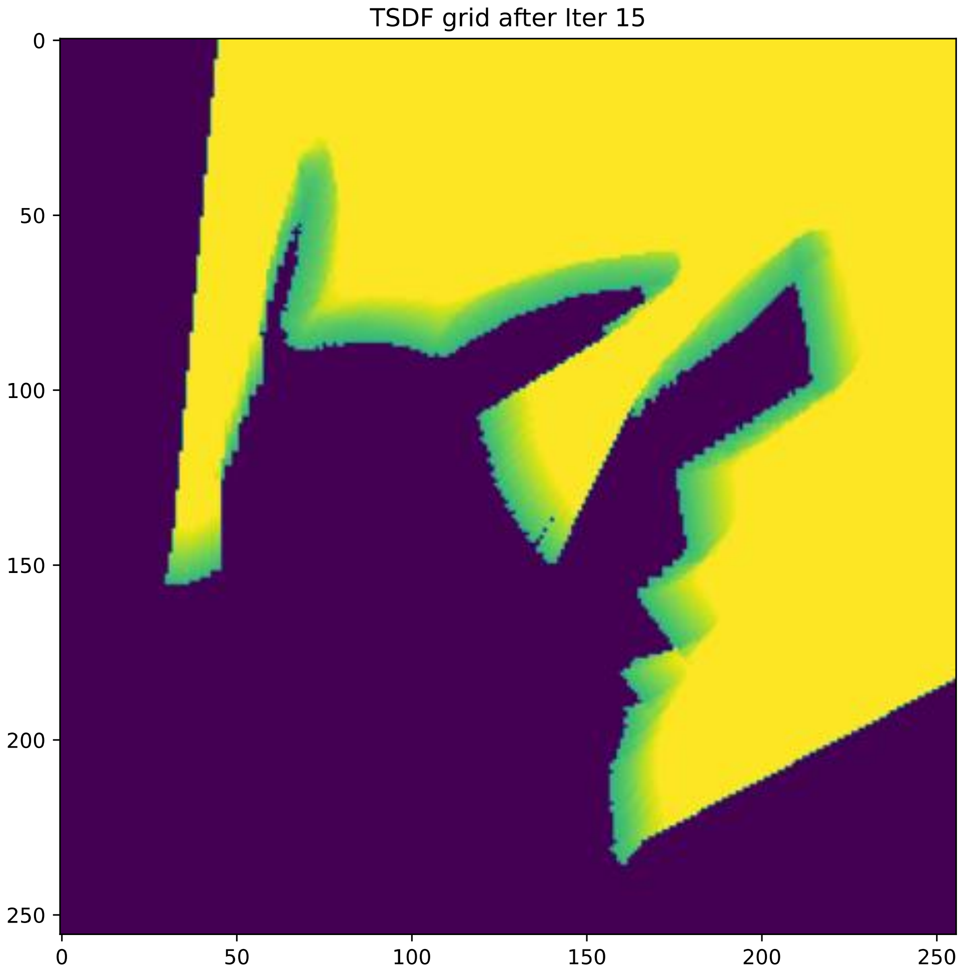




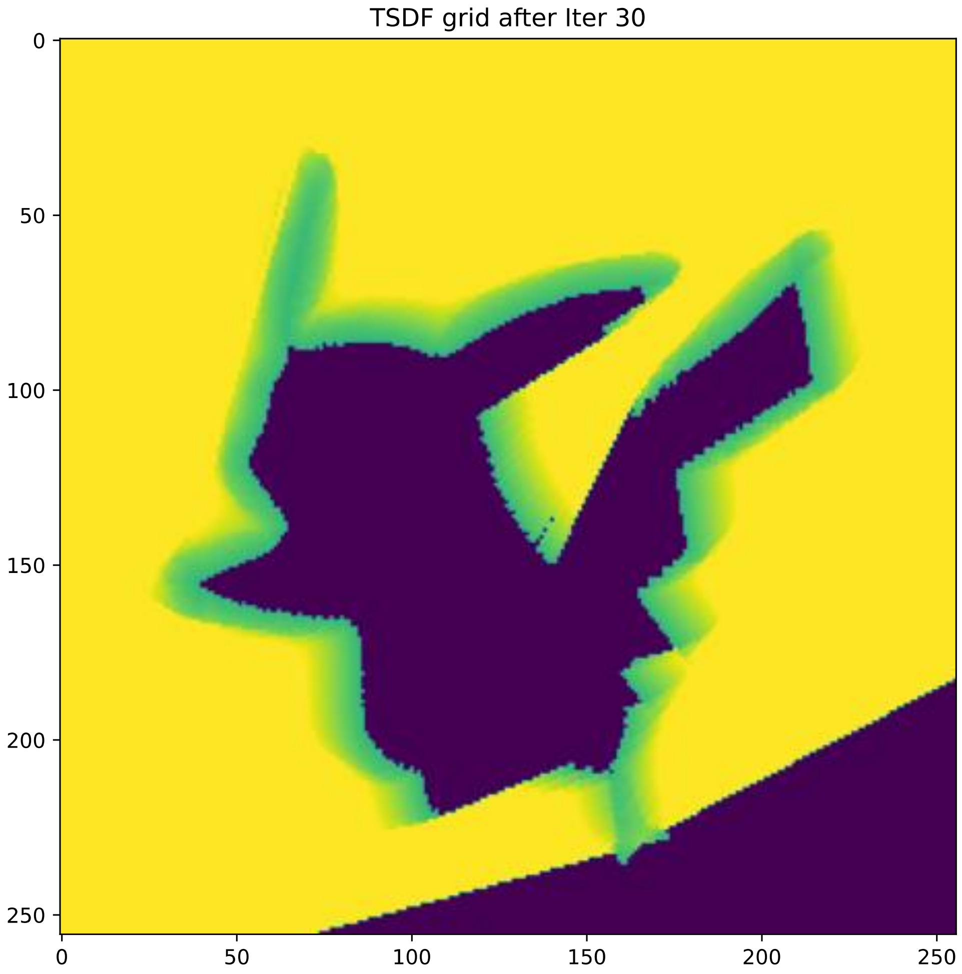
Hacker: Kinect Fusion in a 2D World



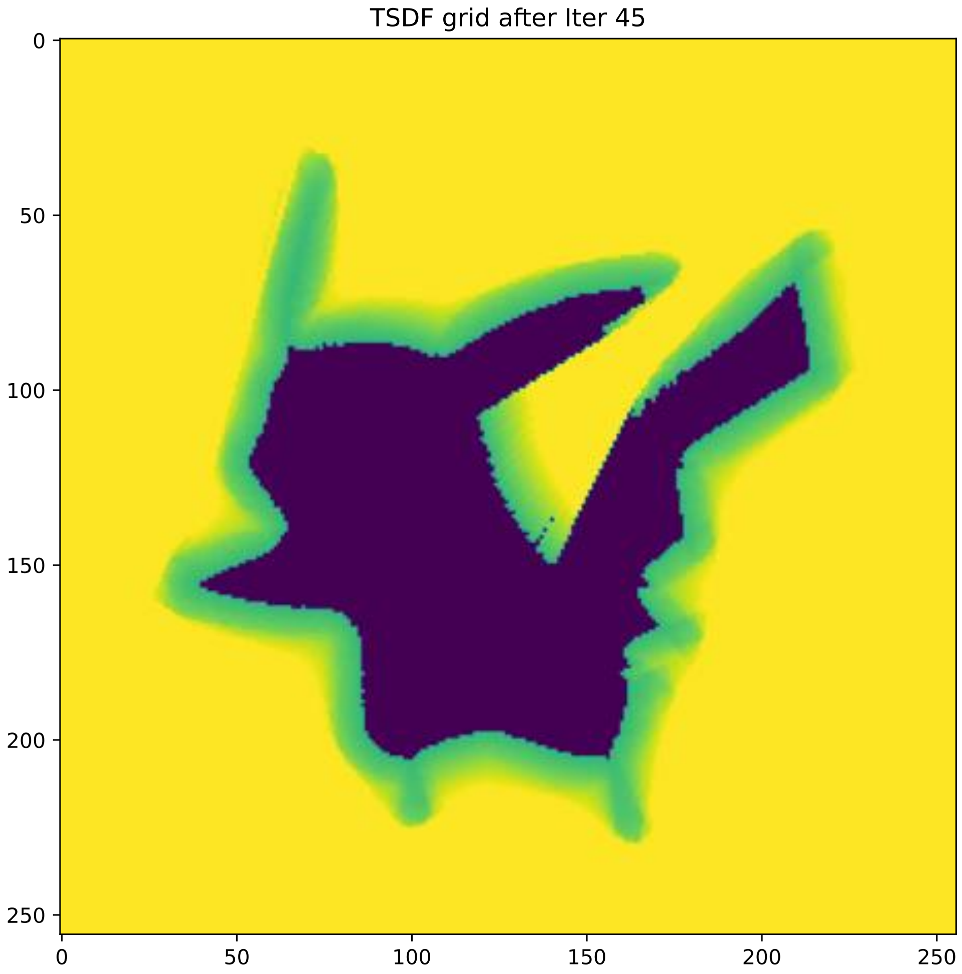
Hacker: Kinect Fusion in a 2D World



Hacker: Kinect Fusion in a 2D World

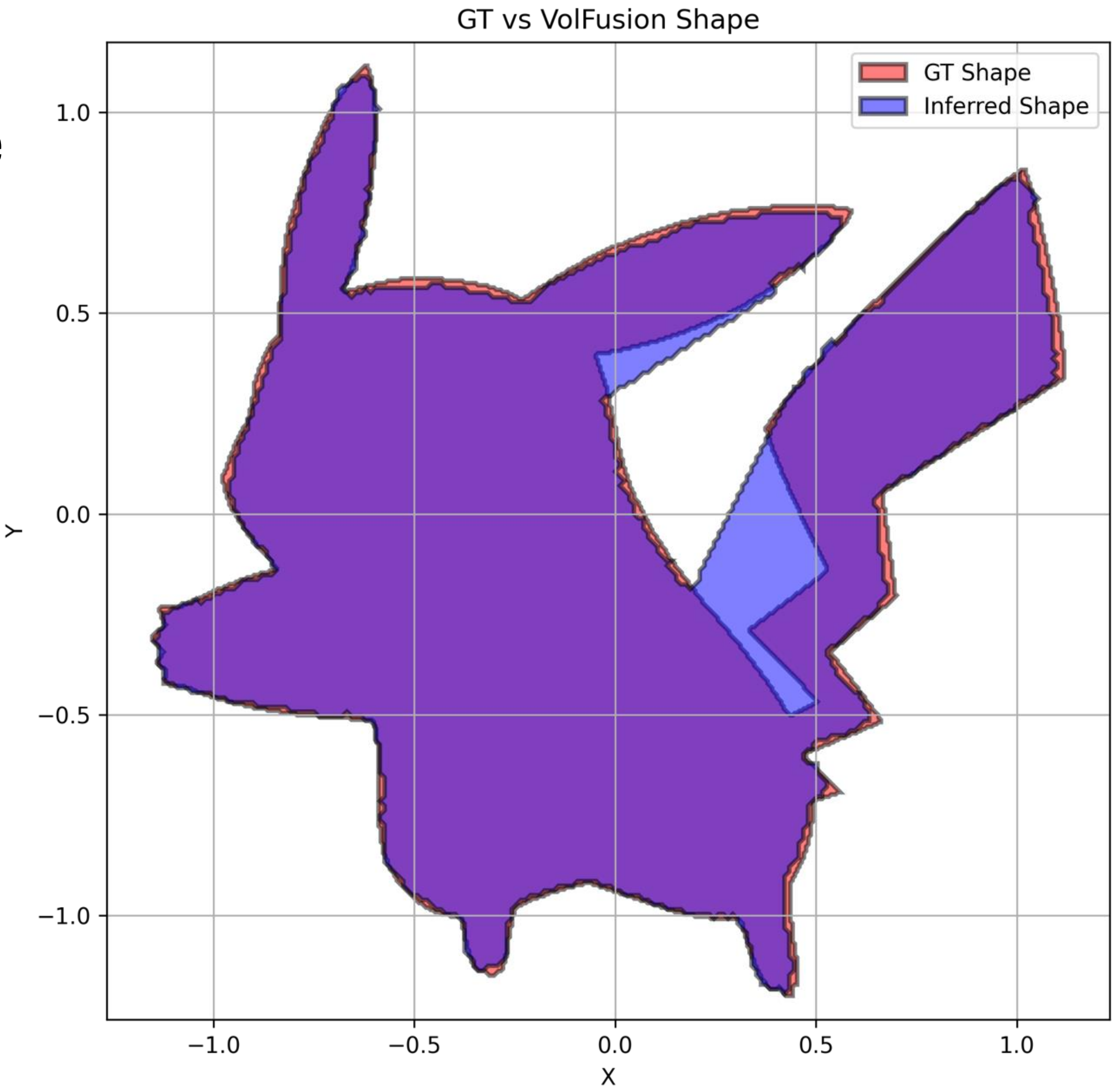


Hacker: Kinect Fusion in a 2D World



Hacker: Kinect Fusion in a 2D World

- Depth sensor has no noise but why it's not perfect?
- Any idea to improve that?



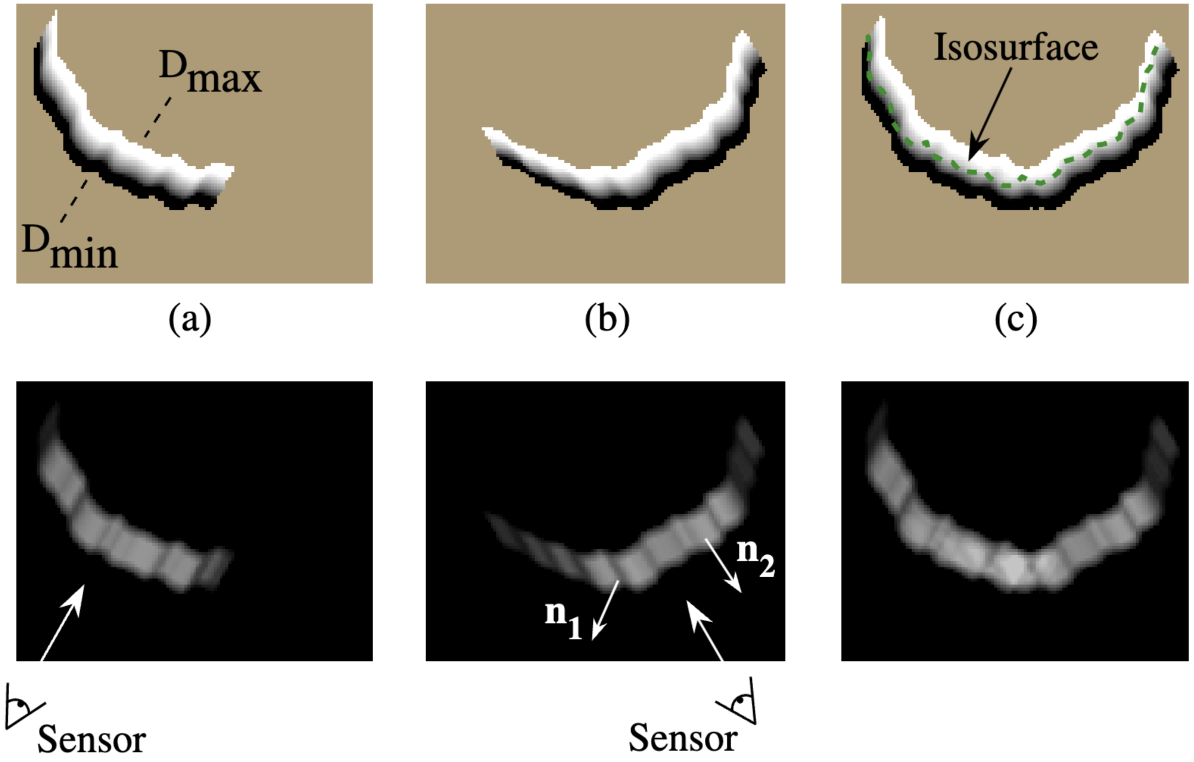
Where I could improve

I kept updating all free space & use a simple weight.

```python
def update_tsdf(tsdf_grid, weight_map, sdf, visibility_mask, trunc_threshold=0.1)
    """
    Update the TSDF grid and the weight map based on the new SDF values.

    Parameters:
    - tsdf_grid: The current TSDF grid.
    - weight_map: The current weight map.
    - sdf: The signed distance function values for the visible points.
    - visibility_mask: Mask of visible points in the grid.
    - trunc_threshold: The truncation threshold for the SDF.

    Returns:
    - Updated TSDF grid and weight map.
    """
    mask = visibility_mask & (sdf > -trunc_threshold)
    tsdf_grid[mask] += sdf[mask]
    weight_map[mask] += 1
    return tsdf_grid, weight_map
```



Points with normal facing sensor should get higher weight, why?

Lessons learned

- Depth sensor has no noise but why it’s not perfect?

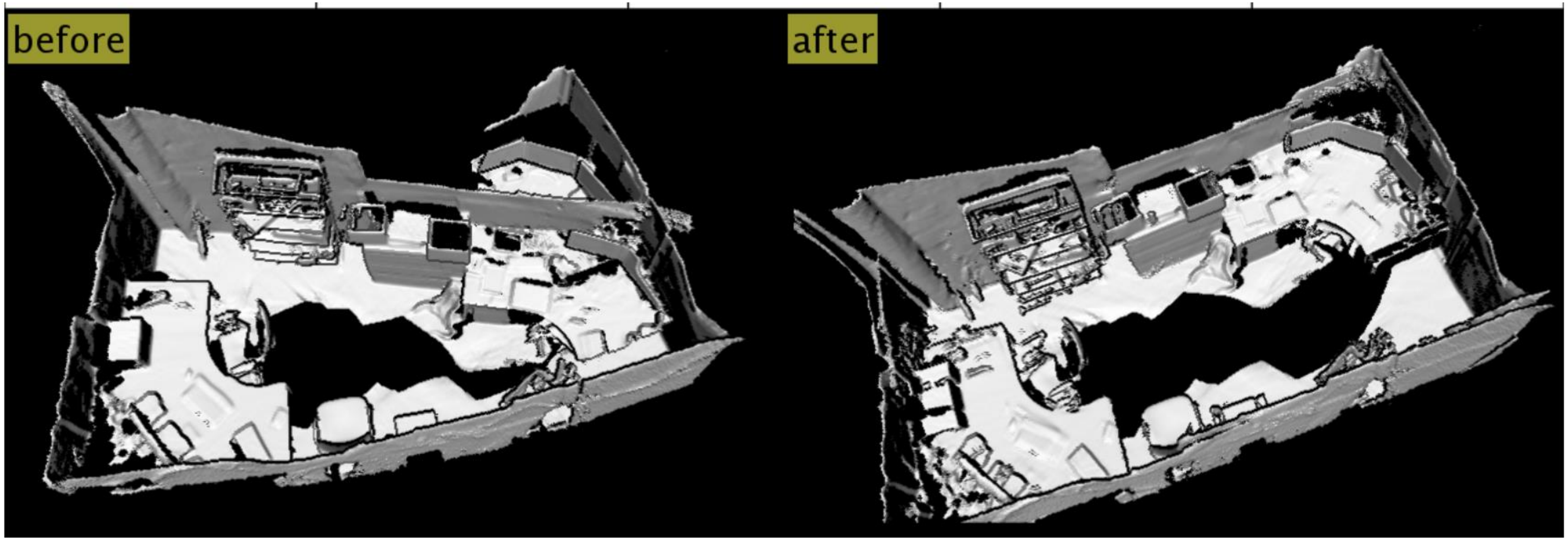
- Any idea to improve that?

Critic

Zhi-Hao Lin as Reviewer #2

Camera drifts with incremental ICP!

- Only works well for small camera motion
- Camera drifts for large planar structure, resulting deformation
- Loop closure requires user manual input – that's not cool!



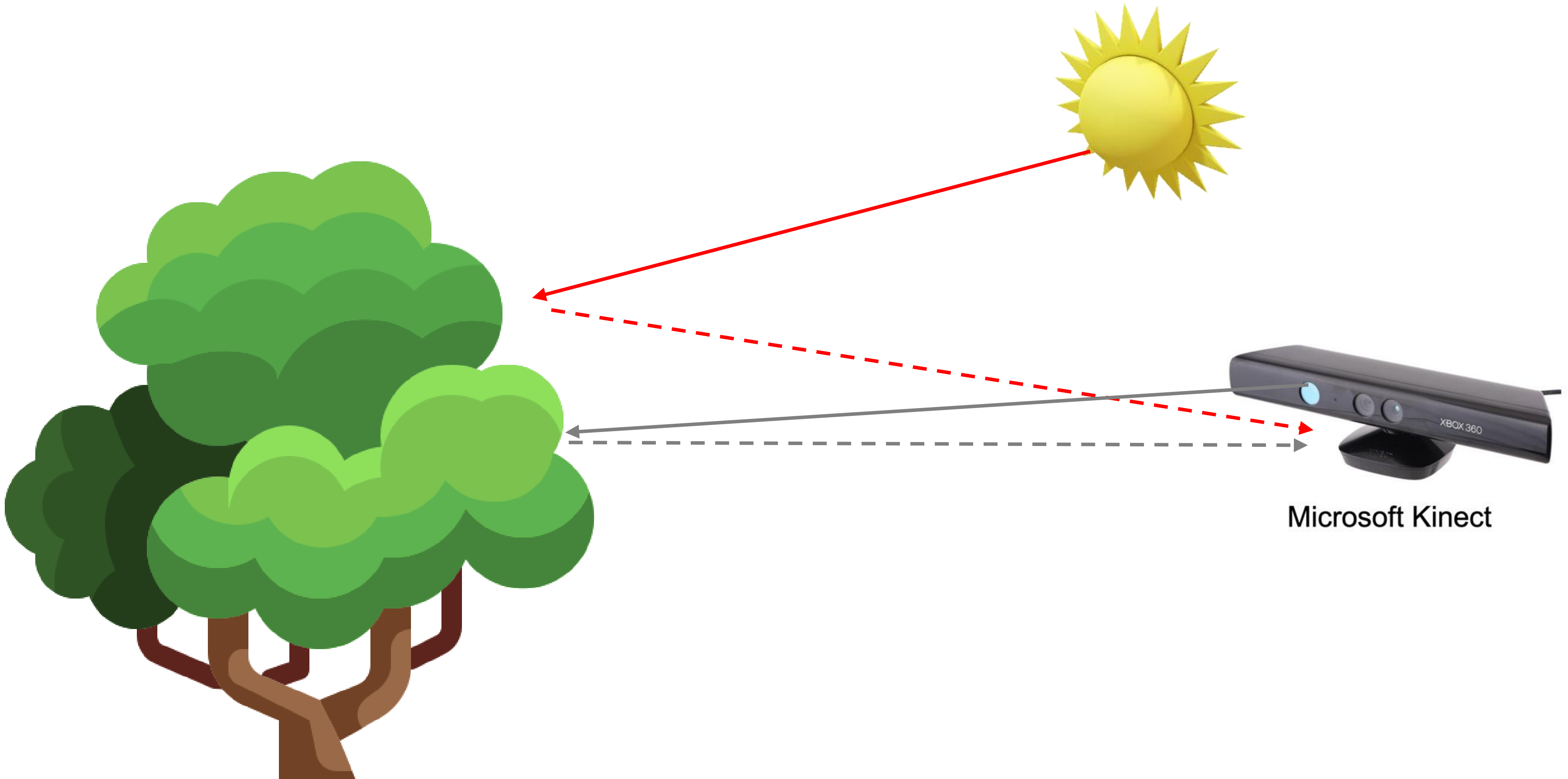
Camera drifts with incremental ICP!

- Only works well for small camera motion
- Camera drifts for large planar structure, resulting deformation
- Loop closure requires user manual input – that’s not cool!

- Does RGB information help?

Only works for indoor!

- Structured light camera (Kinect) doesn't work well in outdoor scenes



We want lidar!

MICROSOFT / TECH

Microsoft kills Kinect again

Image: Microsoft

/ Microsoft will no longer make the Azure Kinect Developer Kit.

By Jay Peters, a news editor who writes about technology, video games, and virtual worlds. He's submitted several accepted emoji proposals to the Unicode Consortium.

Aug 21, 2023, 2:00 PM CDT

40 Comments (40 New)

PRESS RELEASE

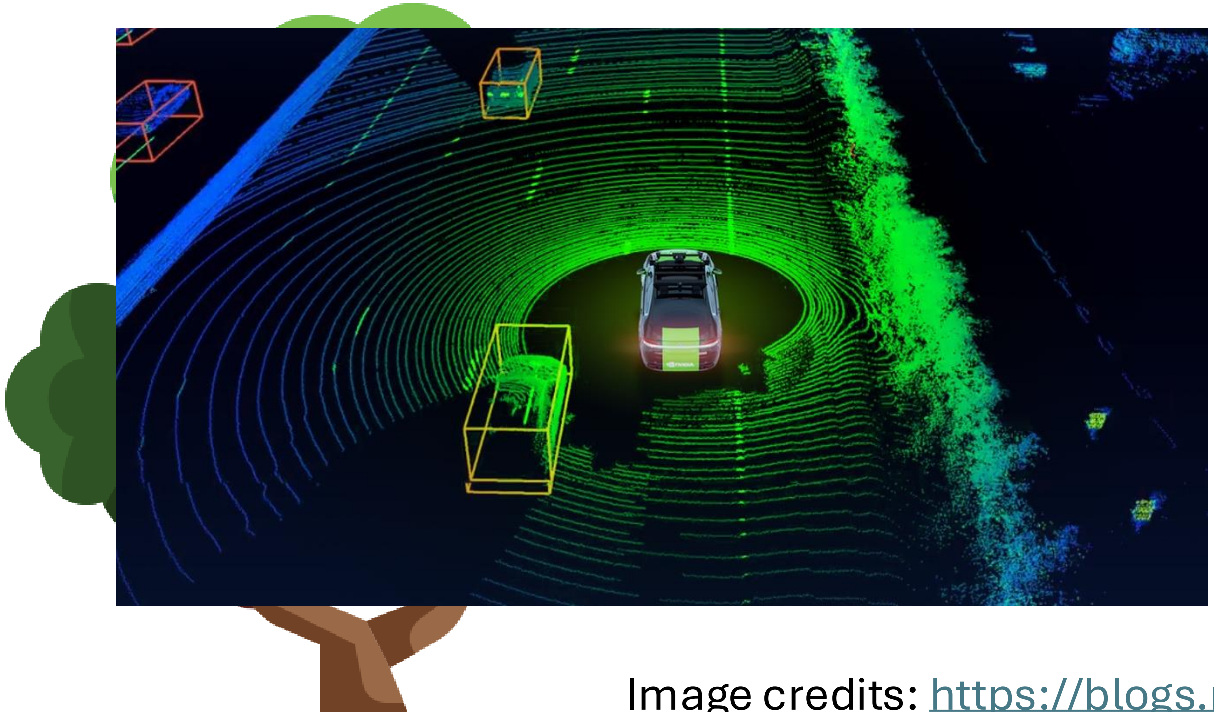
March 18, 2020

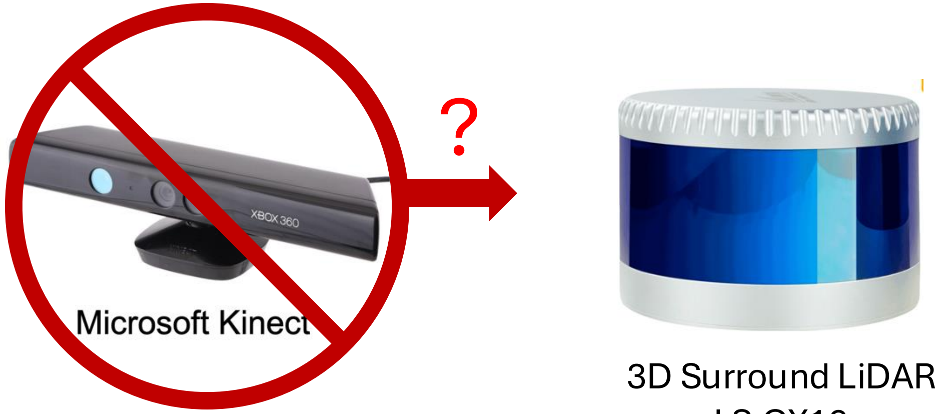
Apple unveils new iPad Pro with breakthrough LiDAR Scanner and brings trackpad support to iPadOS

New Magic Keyboard Designed for iPad Pro Features a Floating Design, Backlit Keyboard and Trackpad, Delivering the Best Typing Experience Ever on iPad

We want lidar!

- Structured light camera (Kinect) doesn't work well in outdoor scenes
- Does the algorithm still work by replacing Kinect with LiDAR?



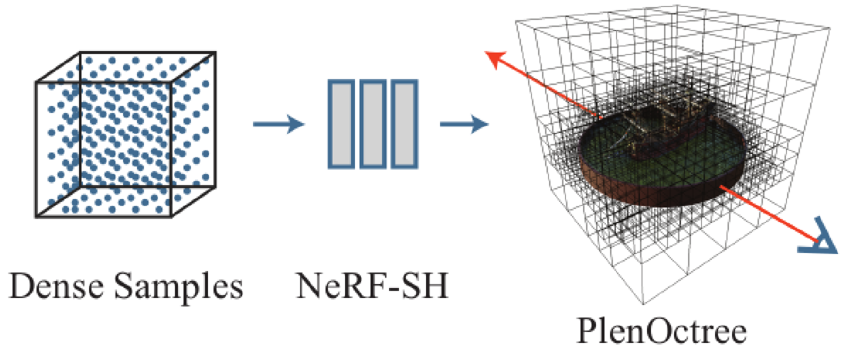


3D Surround LiDAR
LS CX16

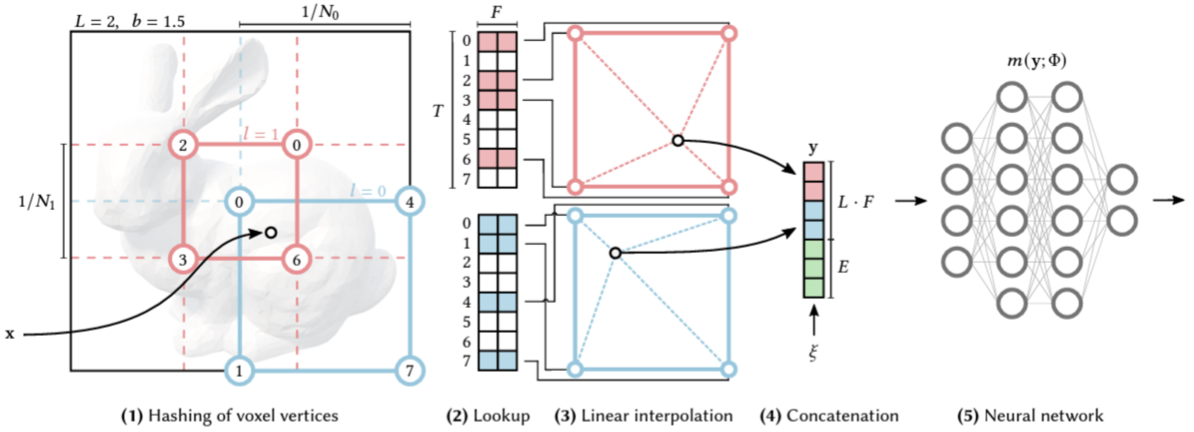
Image credits: https://blogs.nvidia.com/blog/lidar-sensor-nvidia-drive/

Out of memory error!!

- Dense voxel grid is memory-consuming and not scalable
- improve with hierarchical, sparse data structures



PlenOctree: Octree

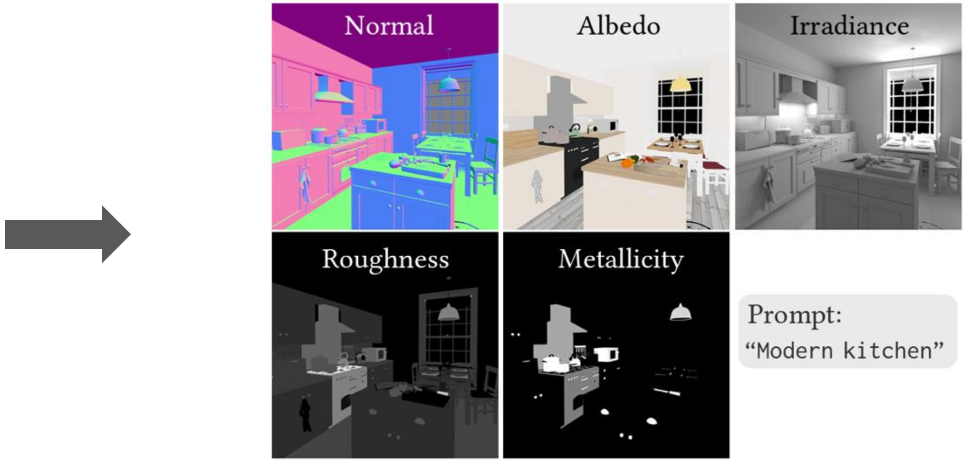


Instant-NGP: Hash encoding

We want more than a surface mesh!

- KinectFusion reconstructs geometry, but some information is missing:

- Transparency
- Material (e.g., roughness, metallic)
- Lighting
- Non-rigid object, motion



RGB↔X: Image decomposition and synthesis using material- and lighting-aware diffusion models

Graduate Student

Albert Zhai

Idea 1: KinectFusion with Learned Depth Prior

Motivation

- Kinect has limited range and has holes in each frame
- Can we leverage learned depth to fill in the gaps and extend the range?

Image credits: KinectFusion, DepthAnything

Idea 1: KinectFusion with Learned Depth Prior

Approach

- Kinect depth can be used to improve learned depth
  - Existing baselines
  - Fit a simple warping per-region from learned to Kinect
  - Fine-tune depth network in the test scene
- Can also consider uncertainty from the fused TSDF

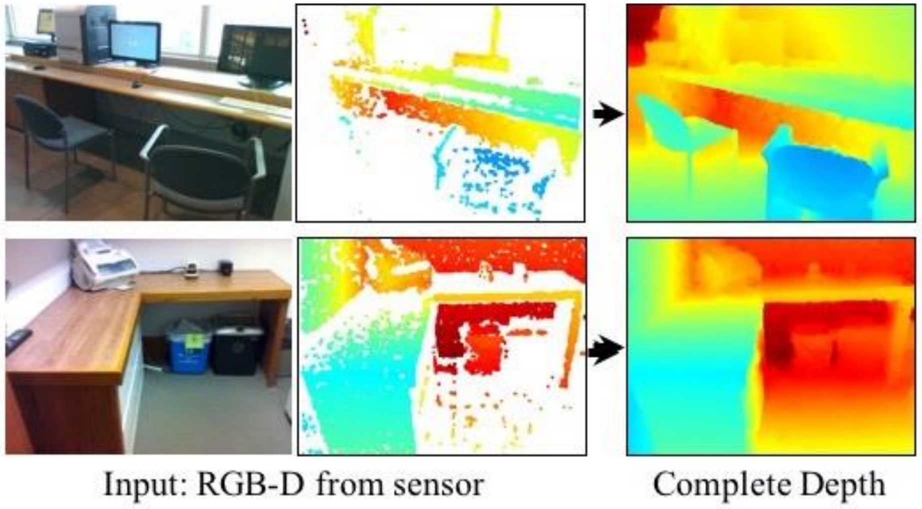
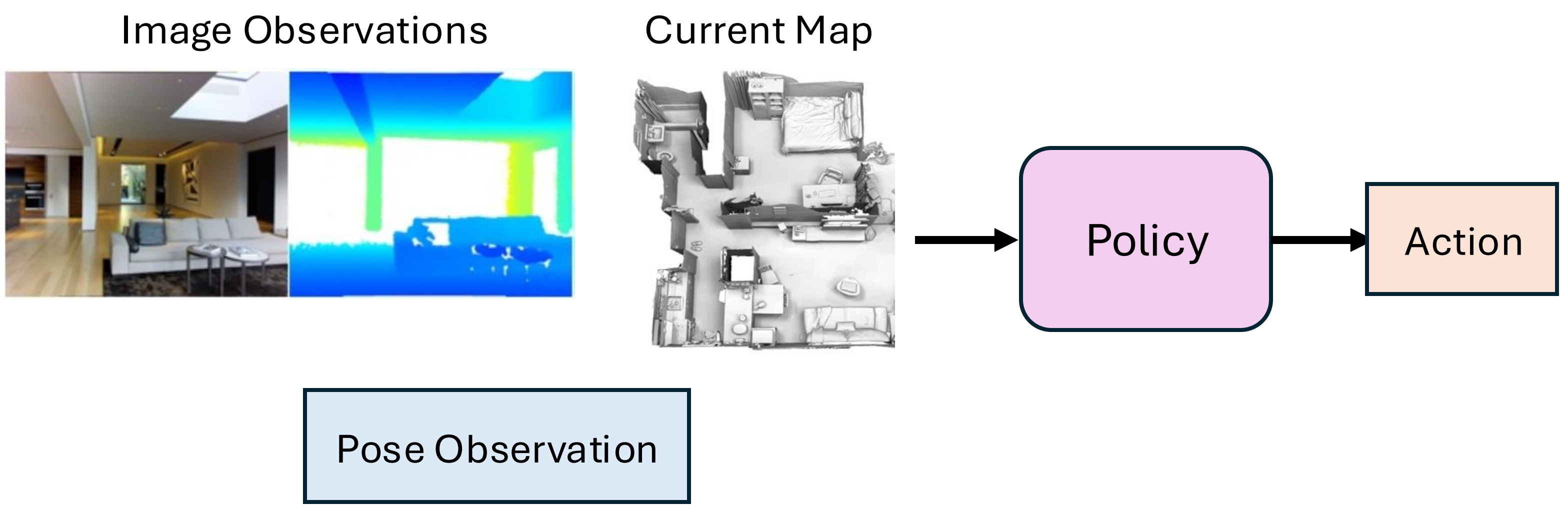


Image credits: DeepCompletion, SuperPrimitives, RCVD

Idea 2: Active Mapping with KinectFusion

- Using KinectFusion, how should we move to build a complete map of a scene as fast as possible?
- Train an agent to do so via reinforcement learning (in sim)
- Design questions: input representation, action space, reward function



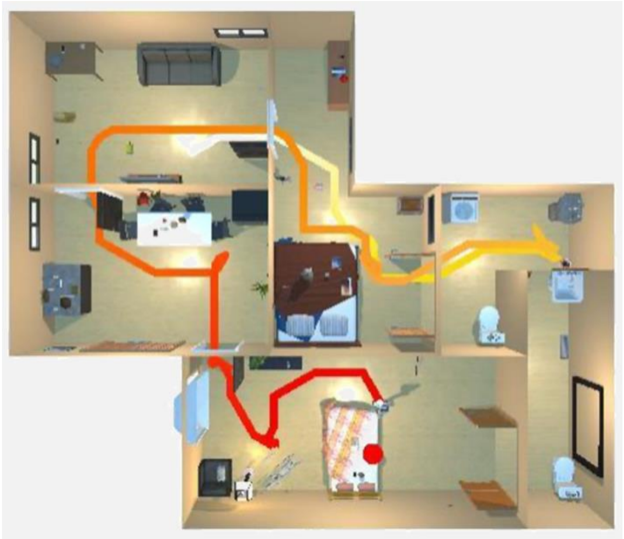


Image credits: DeepCompletion, Qianyi Zhou, SPOC