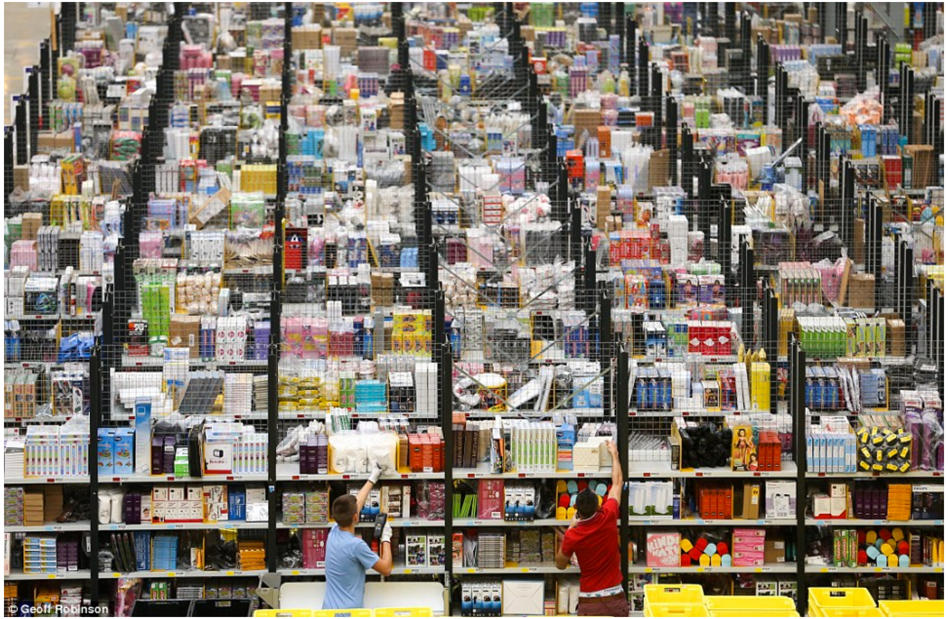
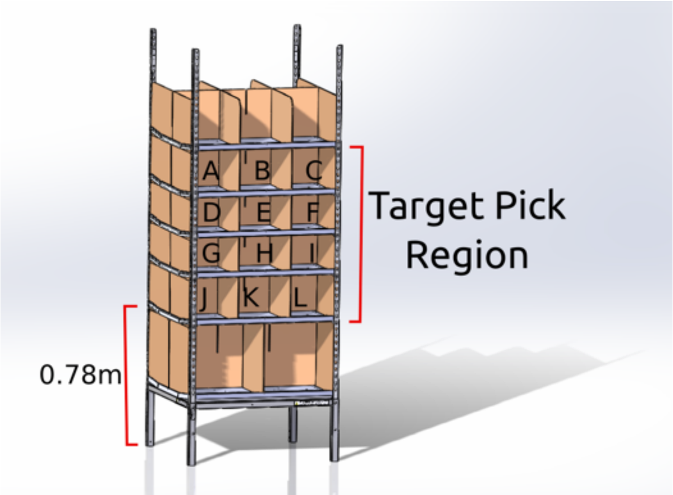
Scalable High-Quality 3D Scanning

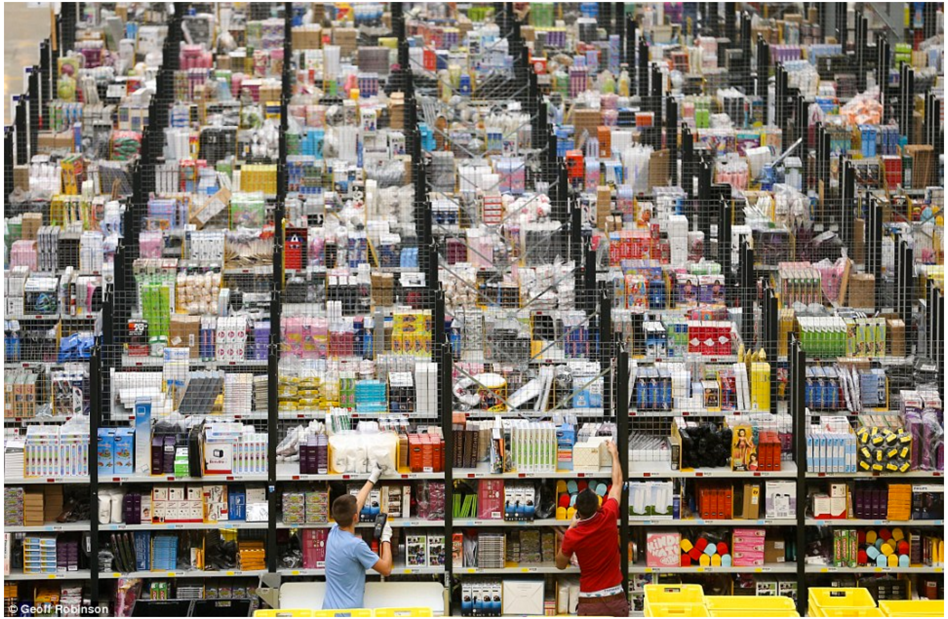
Karthik Narayan

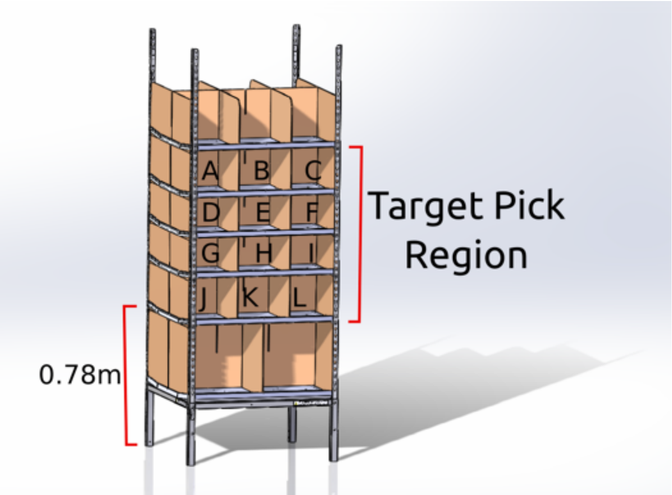
Advisors: Pieter Abbeel, Jitendra Malik


© Geoff Robinson


© Geoff Robinson

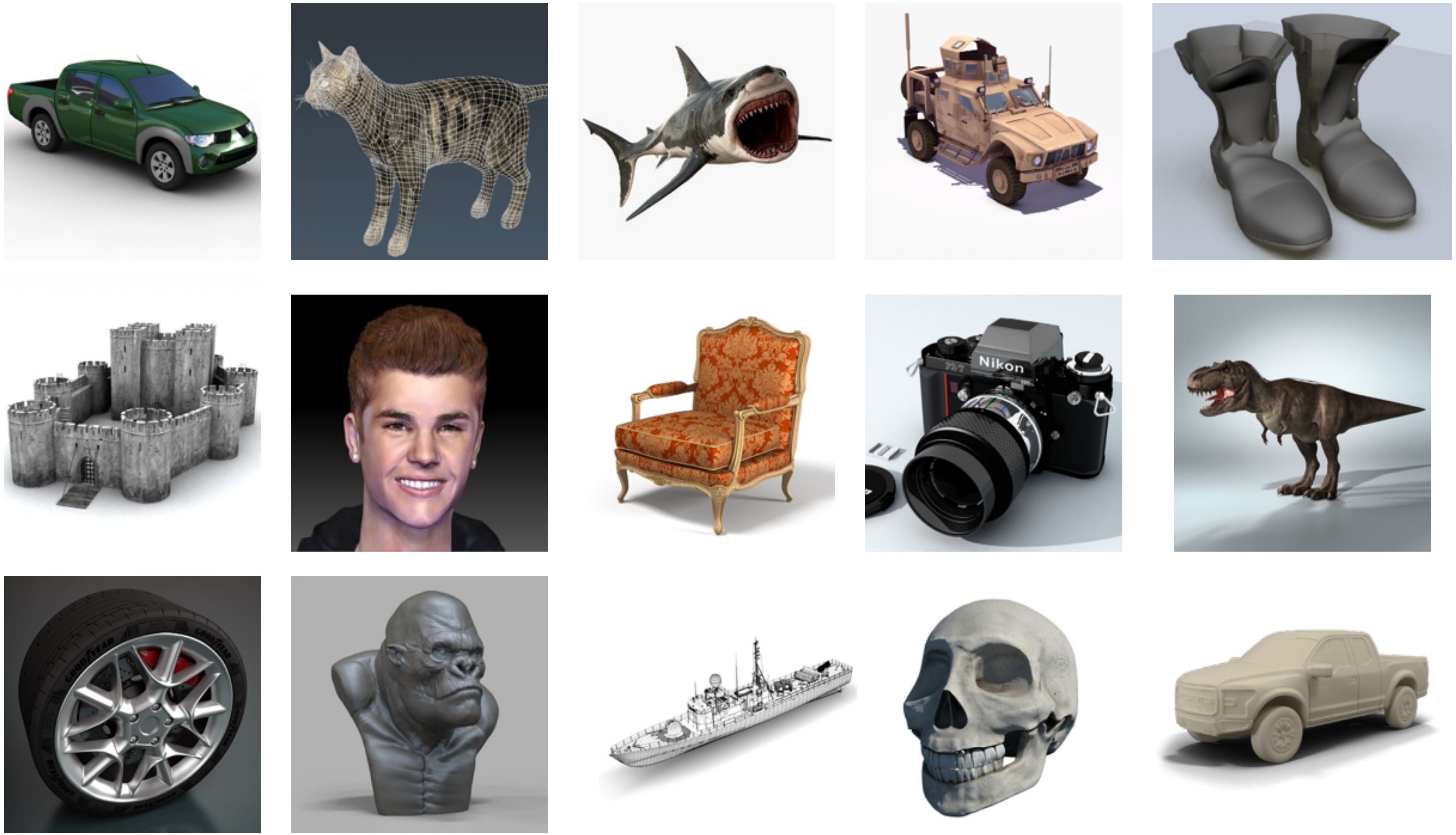

A
B
C
D
E
F
G
H
I
J
K
L
Target Pick Region
0.78m


© Geoff Robinson


A B C
D E F
G H I
J K L
Target Pick Region
0.78m

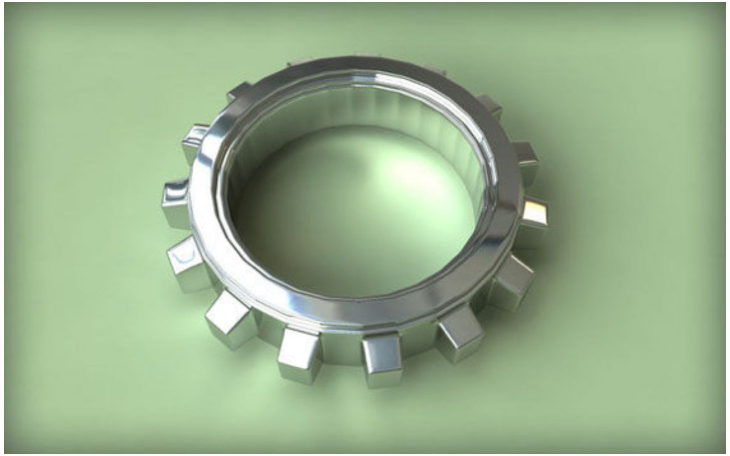

4675
5175
5275
5575
5975
7175
20/09/2013
5200
5227
5539
7121
7130
7134
01/02/2013
20/12/2012
5123
5712/1
6102
4968
5711/1
5940

Scanning lots of objects …



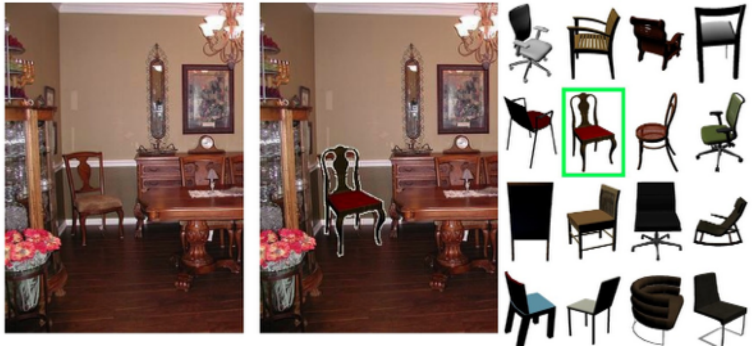
… with high quality.



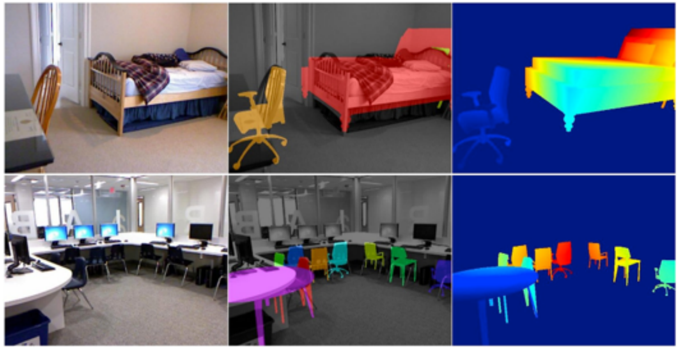





9:51
100



Aubry et. al, CVPR 2014

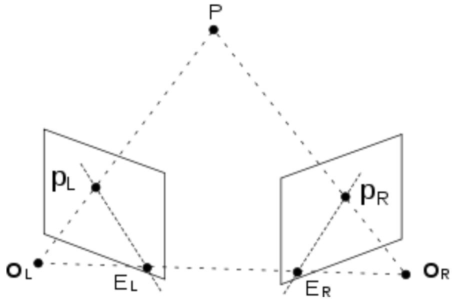


Gupta et. al, CVPR 2015

Xie et. al, IROS 2013

Kholgade et. al, SIGGRAPH 2014

Image-Based Reconstruction



Longuet-Higgins, Nature 1981

Furukawa et. al, CVPR 2007

Seitz et. al, CVPR 2006

Agarwal et. al, ICCV 2009

Active Reconstruction

The Digital Michelangelo Project
($?, but probably a lot)

MakerBot ($650)

NextEngine ($2995)

"Shipping 4 tons of equipment to a foreign country, trucking through narrow streets, and carrying it into historic buildings, was nerve-wracking and expensive."

"During 5 months of scanning, we spent $50K hiring museum guards to watch over us, the statues, and the tourists."

Active Reconstruction

DAVID 3D Scanner ($3275-3395)

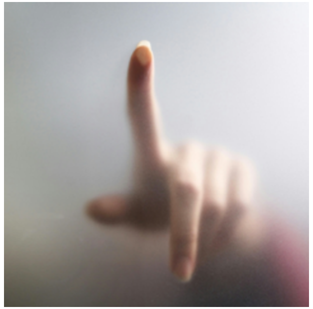
Kinect (<$200)

Artec Eva ($20K)

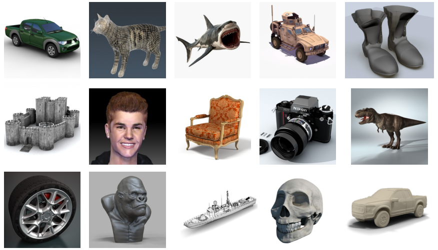
Shortcomings

Cost

Object Size



Translucencies



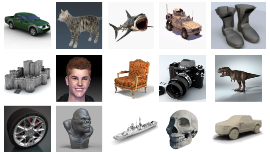
Scanning Many Objects

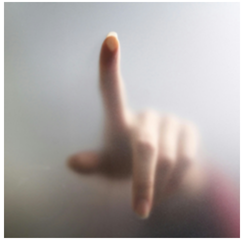
Color Extraction

Dark Objects

Scalable High-quality 3D Scanning

- Scalability
  - How do we scan many objects?
  - How do we keep device costs low?

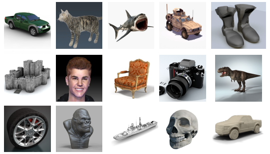



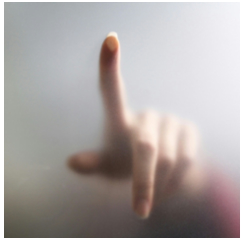


- High quality
  - How do we extract high quality shape models?
  - How do we extract high quality color models?


Scalable High-quality 3D Scanning

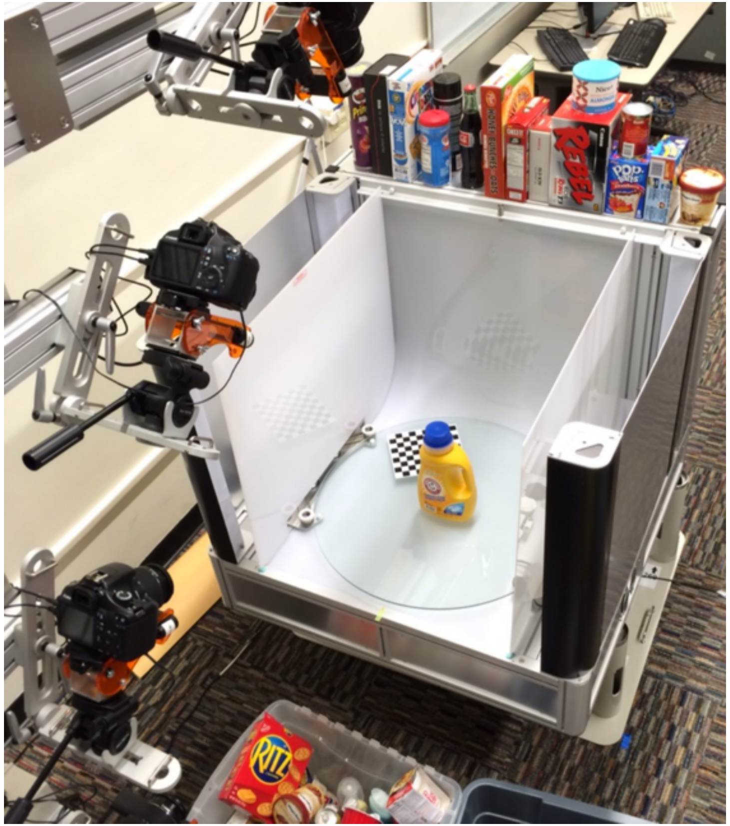
- **Scalability**
  - How do we scan many objects?
  - How do we keep device costs low?

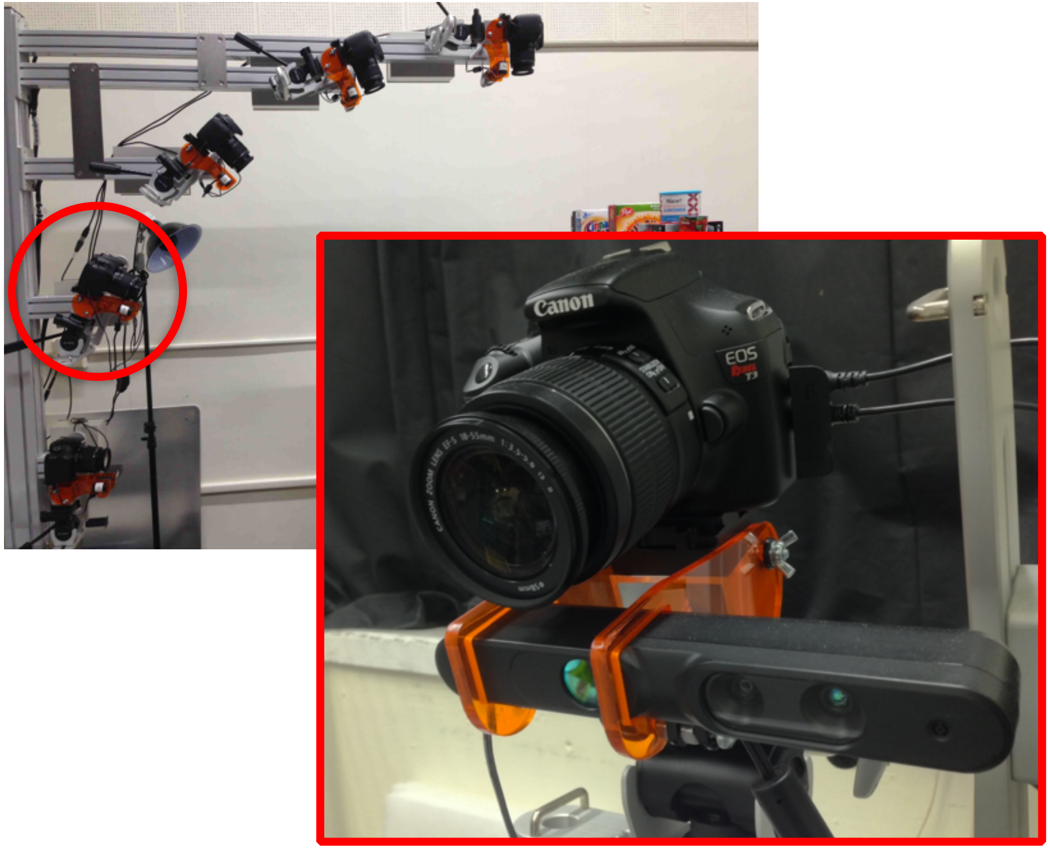




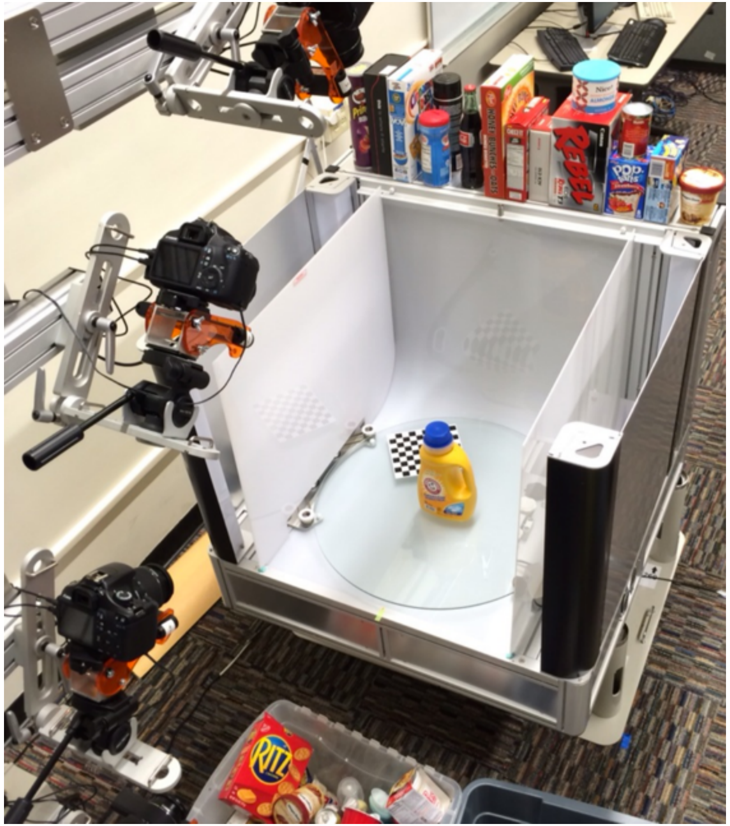
- High quality
  - How do we extract high quality shape models?
  - How do we extract high quality color models?

Big Berkeley Instance Recognition Dataset (BigBIRD)





BigBIRD

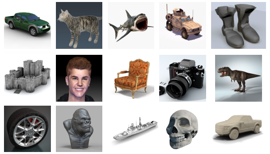


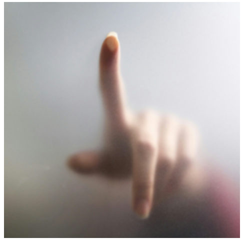
- 600 12 MP DSLR images
- 600 Kinect RGB+D images
- All images are calibrated via two bundle adjustment procedures
- Scanning 1 object involves pushing a button after calibration.

BigBIRD: A Large-Scale 3D Database of Object Instances,
A. Singh, J. Sha, K. Narayan, T. Achim, P. Abbeel.
ICRA 2014.

Scalable High-quality 3D Scanning

- Scalability
  - How do we scan many objects?
  - How do we keep device costs low?
- **High quality**
  - **How do we extract high quality shape models?**
  - **How do we extract high quality color models?**

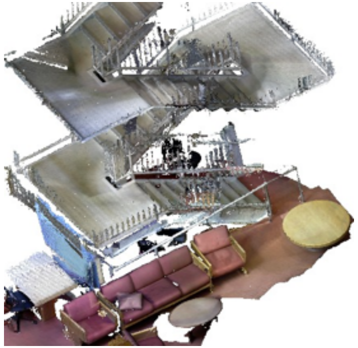


3D Scanning - Components

- Shape estimation

Newcombe et. al. ISMAR 2011

Zhou et. al. CVPR 2014



Whelan et. al. RSS-W 2012

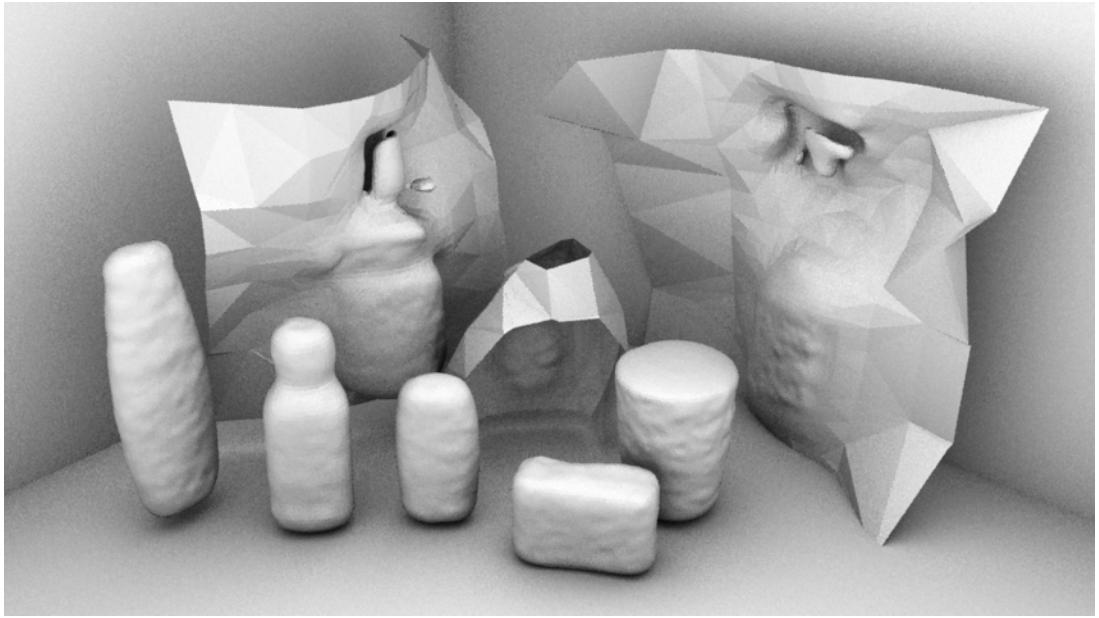
- Color estimation

Hernandez, Thesis 2004

Zhou et. al. ACM TOG 2014

Rusu, ICRA 2011

Shape Estimation



Poisson reconstruction on all Kinect clouds

Poisson Surface Reconstruction. M. Kazhdan, M. Bolitho, and H. Hoppe. ESGP 2006.

Shape Estimation

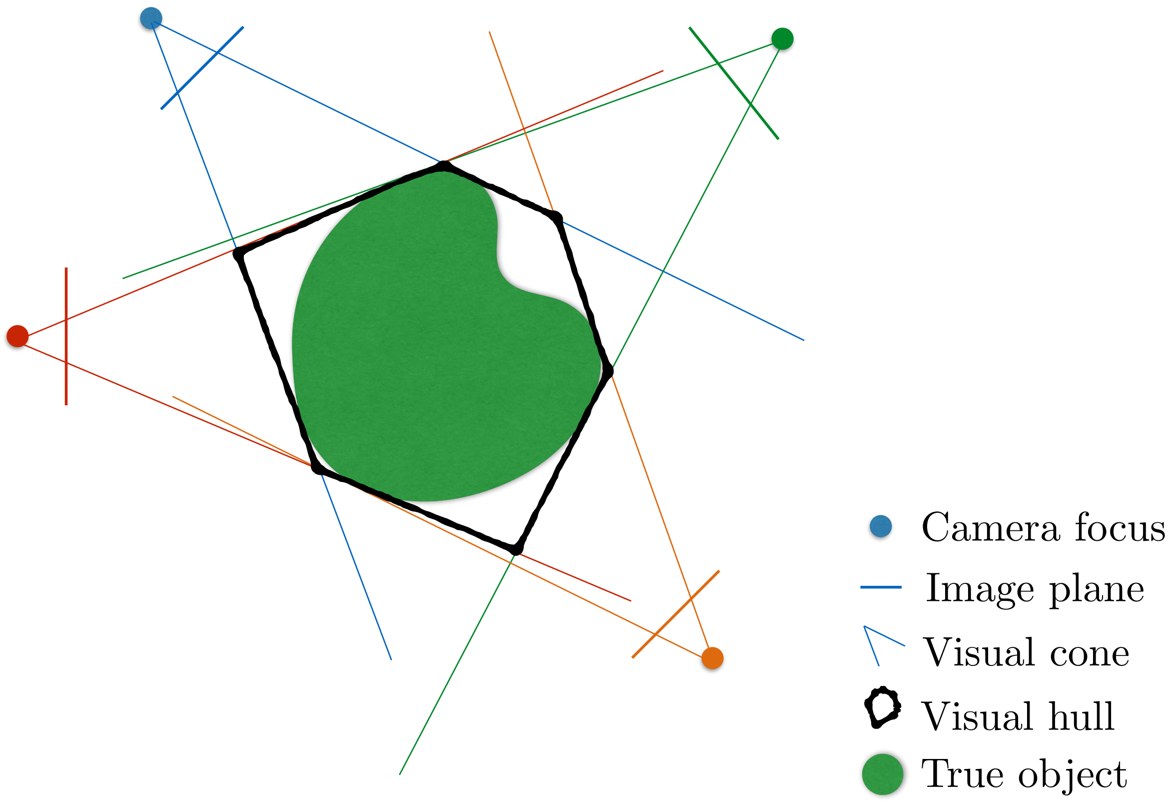
KinectFusion-based approach.

Shape Estimation

New approach. Fuse RGB + D.

Range Sensor and Silhouette Fusion for High-Quality 3D Scanning. K. Narayan, J. Sha, A. Singh, P. Abbeel. ICRA 2015.

Visual Hull

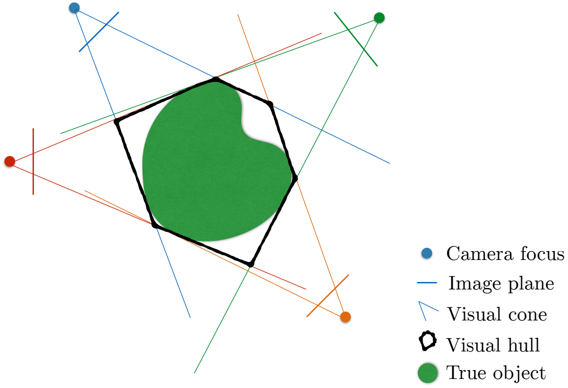

Camera focus
Image plane
Visual cone
Visual hull
True object

KinectFusion

- Real-time
- Fails in reconstructing objects with transparencies, specularities, and thin features.
- Details are often smoothed away.

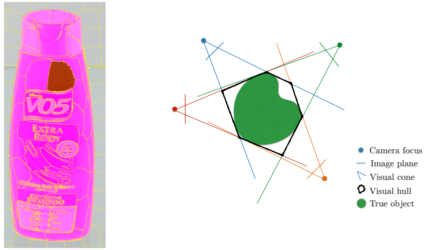
Insight

	KinectFusion	Visual Hull
Concavities	✓	✗
Transparencies	✗	✓
Surface Detail	✓	✓



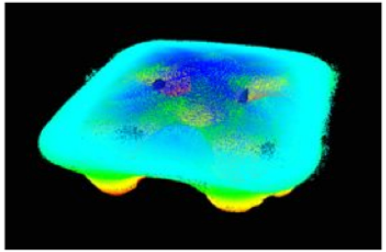
Refine the depth maps to integrate information provided by both of these modalities.

Our Approach

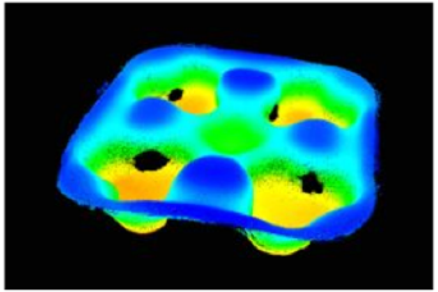


Segmentation + Visual Hull

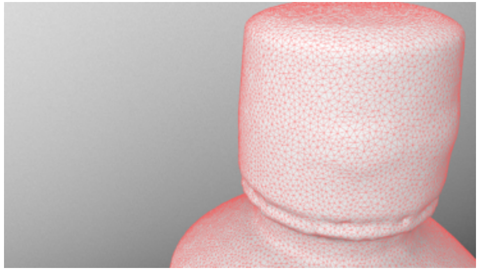
KinectFusion



Compute Fused Cloud



Postprocess Cloud



Mesh Generation

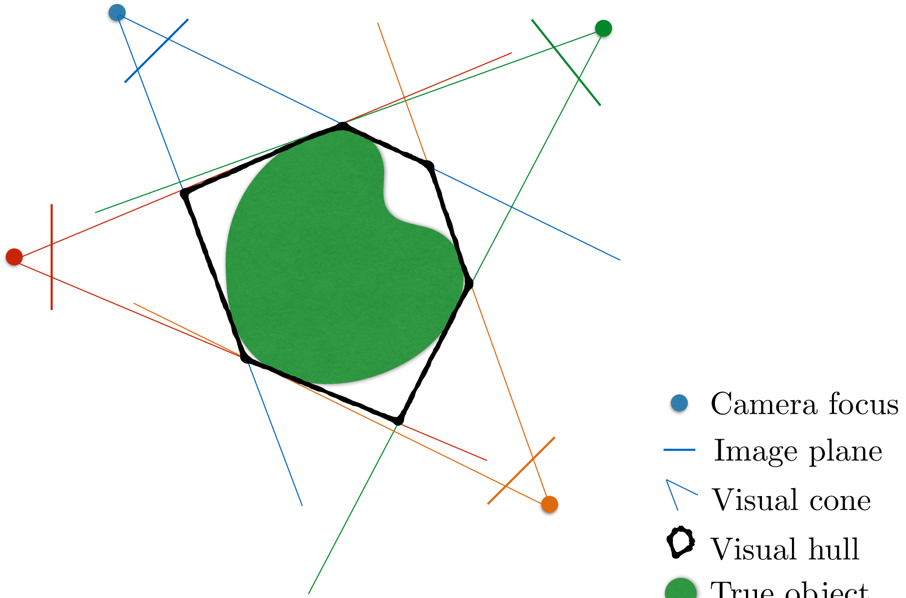
Segmentation

- Manual SLIC segmentation on 5 objects
- Build k-means background models
- Pixels farther than a threshold belong to the object.
- Retain super pixels with > 30% coverage.

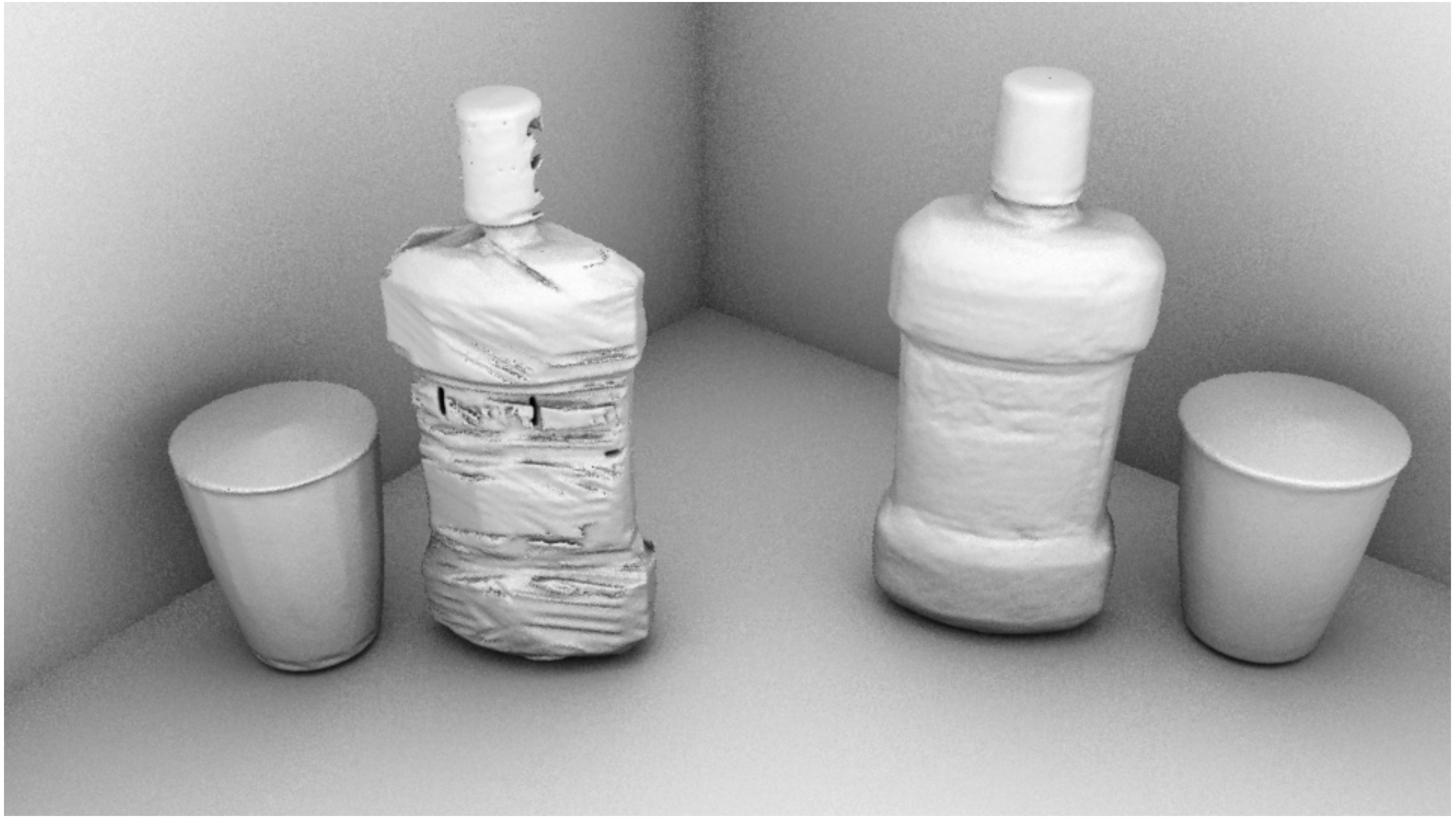
(a) Interactive Segmentation (b) A&H Detergent (c) VO5 Shampoo (d) Softsoap Handsoap

Visual Hull

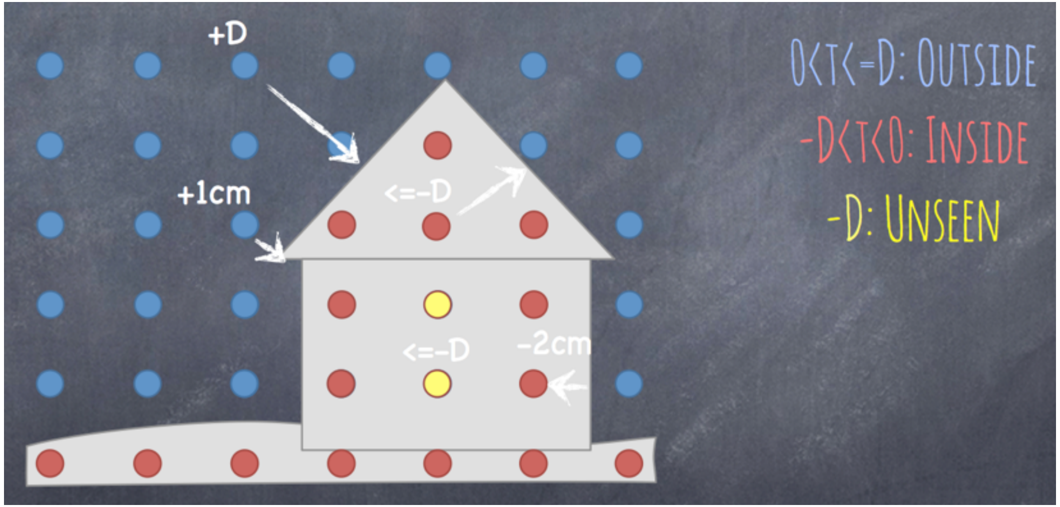
- Construct an implicit function F(x). x is a 3D point.
  - F(x) = 1, x is "inside the object"
  - F(x) = 0, x is "outside the object"
- x is "inside the object" if it projects into 1 - e of the segmentation masks, a.k.a. silhouettes.
- Bloomenthal polygonization



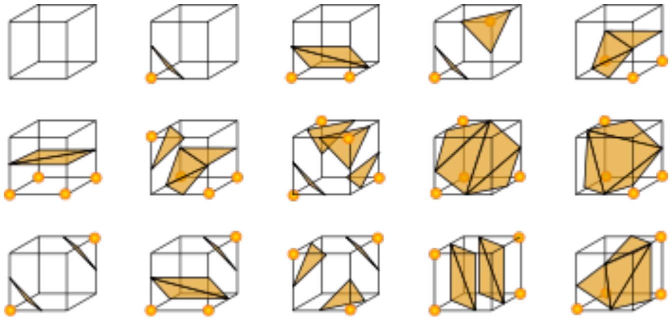
Visual Hull



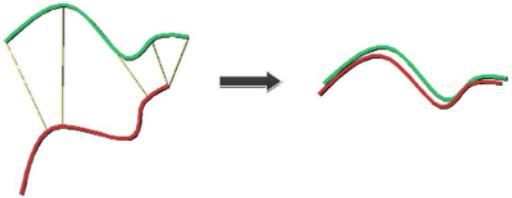
KinectFusion



Truncated signed distance function (TSDF)

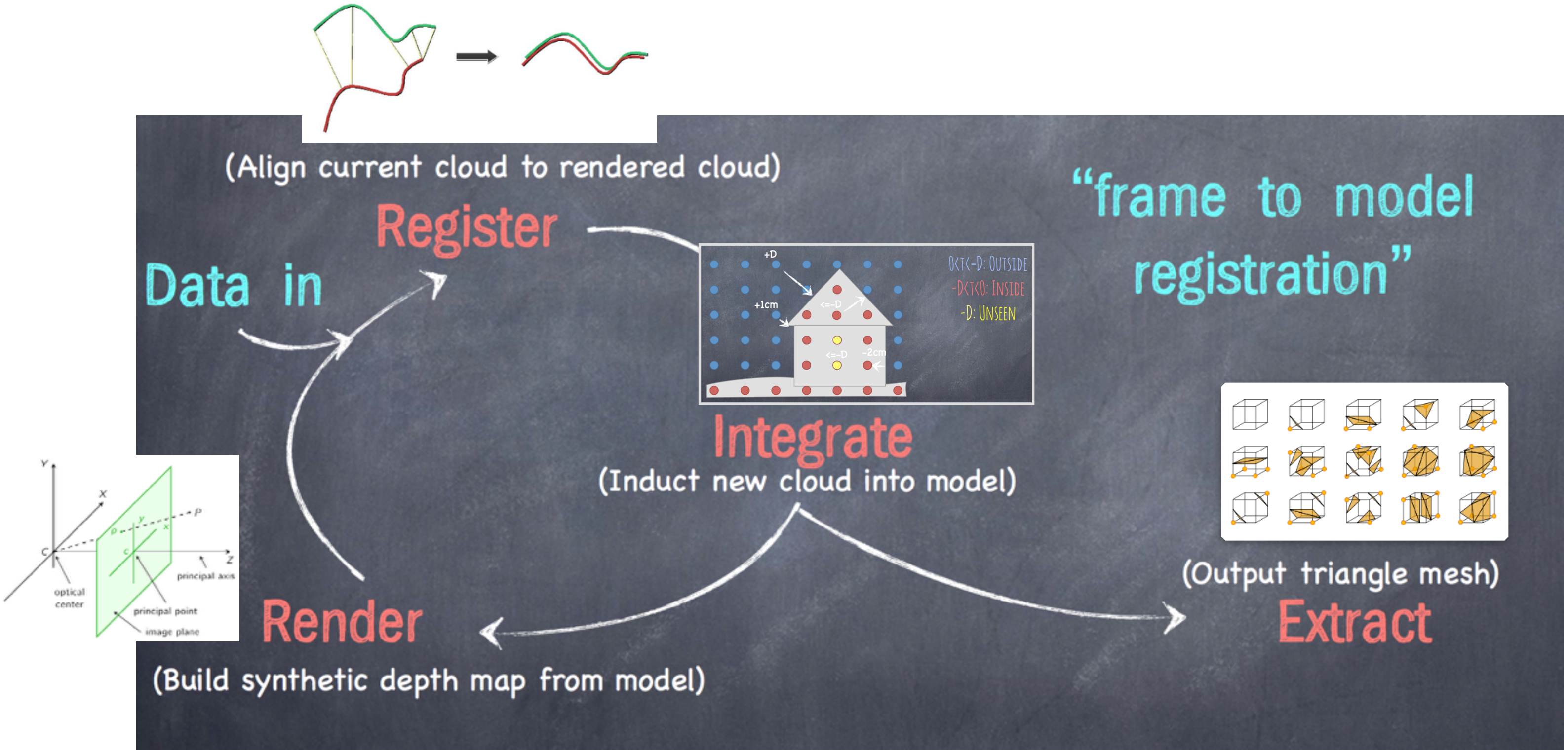


Marching Cubes



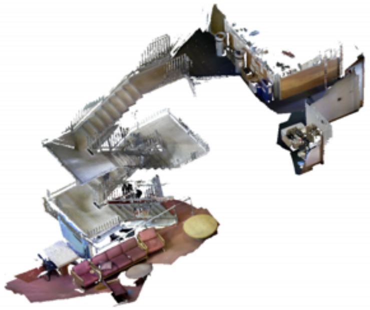
Iterative Closest Point

KinectFusion

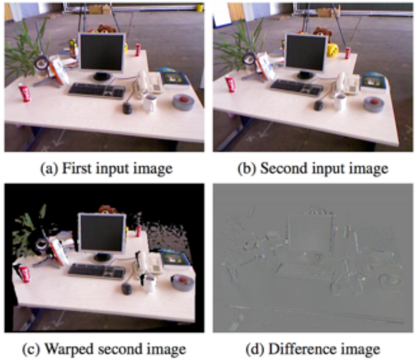


KinectFusion Variants

Zhou et. al. CVPR 2014



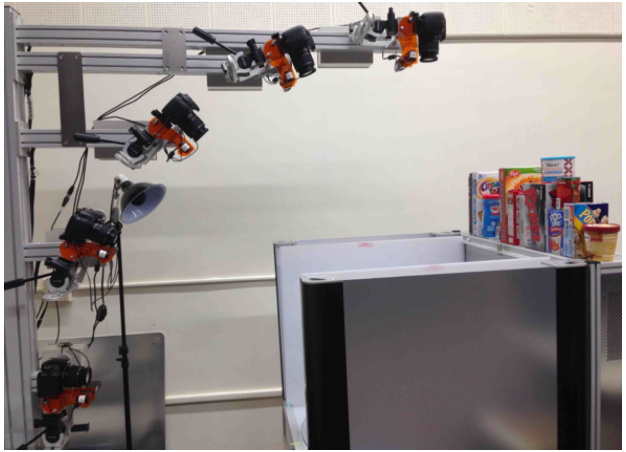
Whelan et. al. ICRA 2013



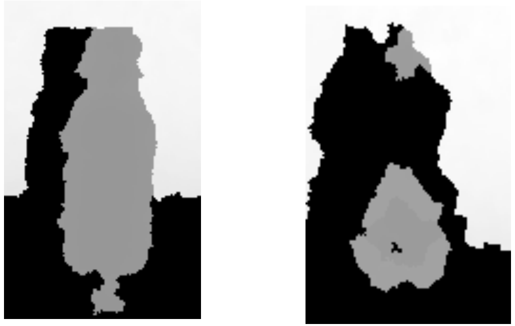
Steinbrucker et. al. ICCV 2011

Zhou et. al. SIGGRAPH 2014

KinectFusion Challenges



Cameras are far apart



Snippets of (incomplete) depth data

Construct TSDF via BigBIRD calibration

KinectFusion

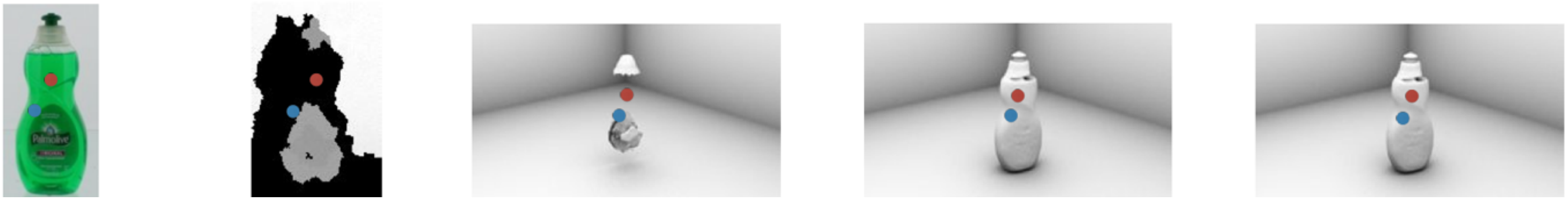
KinectFusion-based approach.

Depth Refinement

- Idea:
  - Consider a single depth map associated with camera c and angle a.
  - At a pixel (i, j), how do we combine the visual hull and KinectFusion depths?

Missing Depth Values

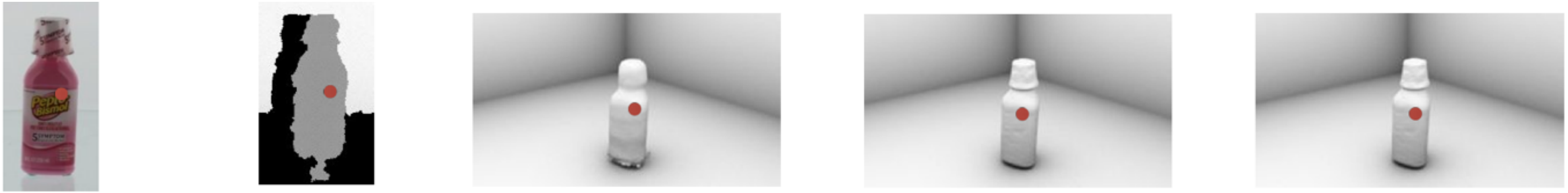
- This happens on/near transparent regions.
- Resort to using the visual hull's depth.



(a) Object color images (b) Raw depth maps (c) KinectFusion meshes (d) Soft visual hull meshes (e) Our method

KinectFusion and Visual Hull Agree

- This happens on “reliable” surfaces.
- Resort to using the visual hull’s depth map for finer surface details.



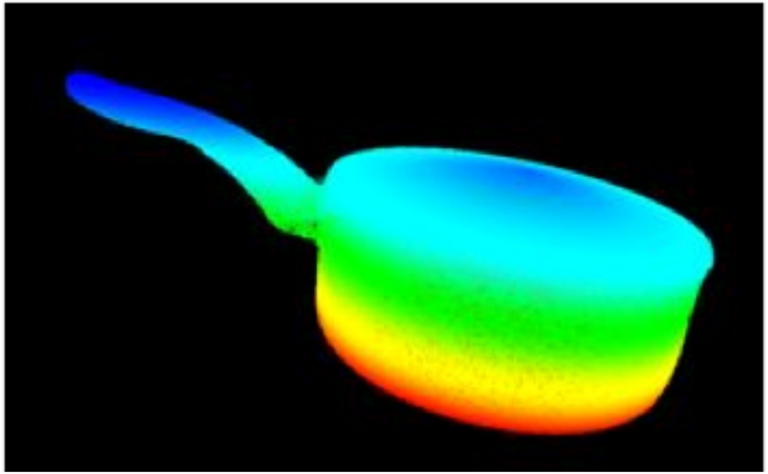
(a) Object color images (b) Raw depth maps (c) KinectFusion meshes (d) Soft visual hull meshes (e) Our method

KinectFusion and Visual Hull Disagree

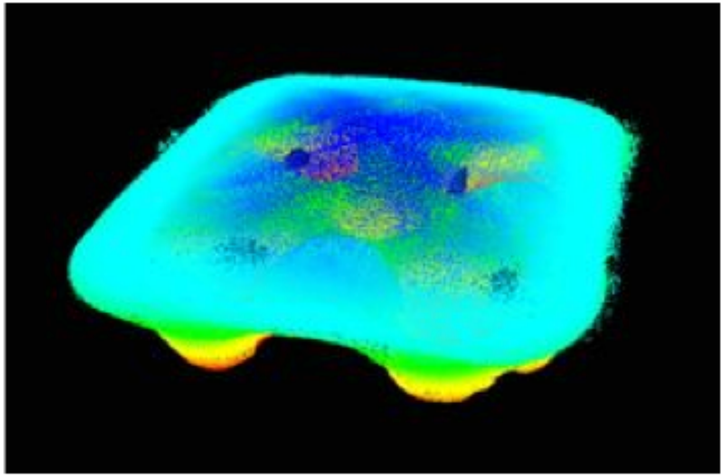
- This happens at concavities.
- Resort to using the KinectFusion's depth map.

(a) Object color images (b) Raw depth maps (c) KinectFusion meshes (d) Soft visual hull meshes (e) Our method

Lots of Hallucinated Points

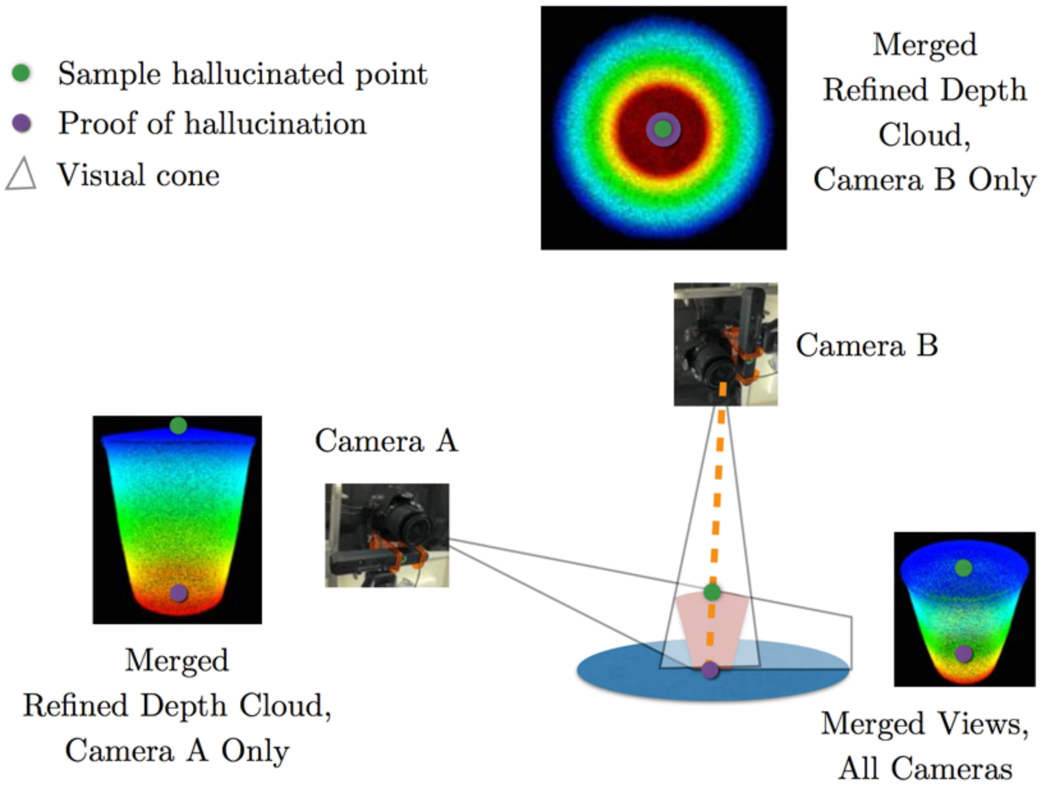


(a) Pot, with hallucinated points

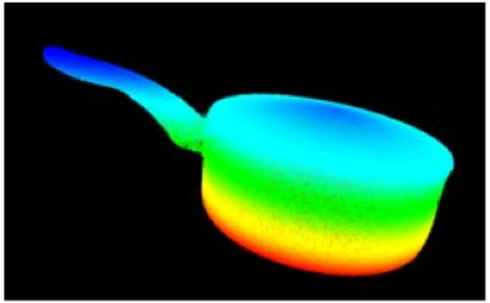


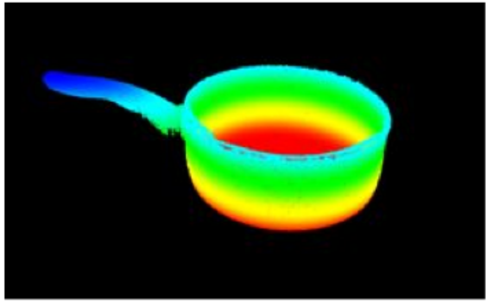
(b) Cup holder, with hallucinated points

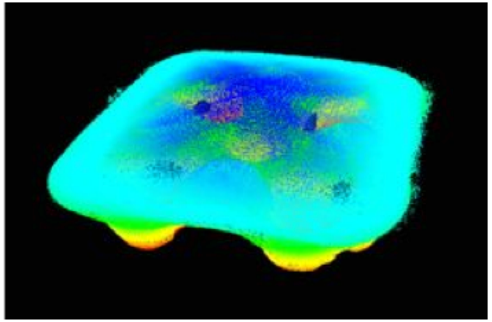
What Causes Hallucinations?

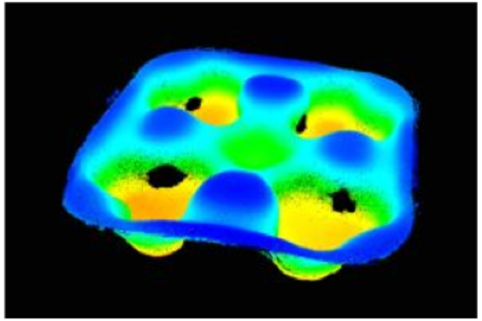


Hallucinations: Before and After


(a) Pot, with hallucinated points

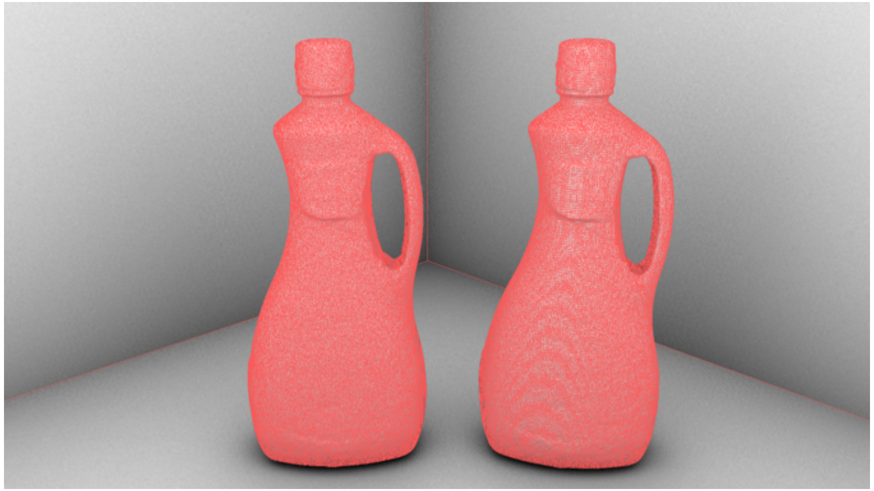

(b) Pot, hallucinated points removed


(c) Cup holder, with hallucinated points

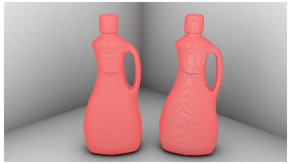

(d) Cup holder, hallucinated points removed

Final Mesh Creation

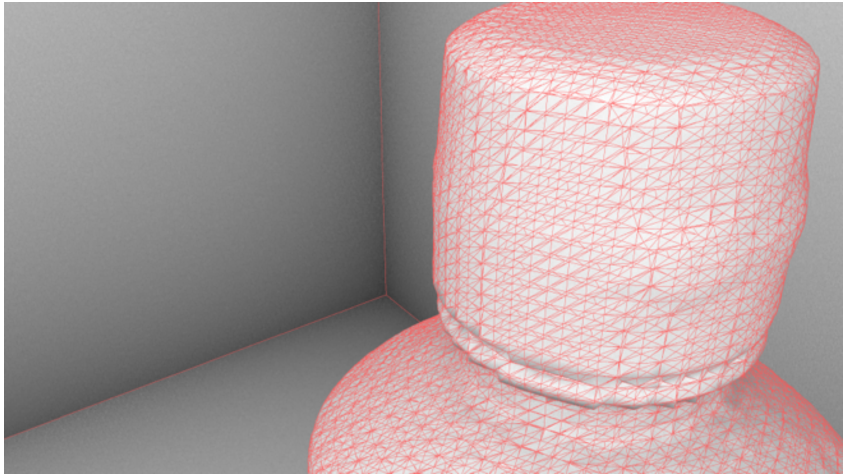
- To construct a mesh, define a function F(x). x is a 3D point:
  - F(x) = 1, x lies within 1 mm of a point in the de-hallucinated cloud AND lies within 1 - e of the silhouettes.
  - F(x) = 0, otherwise



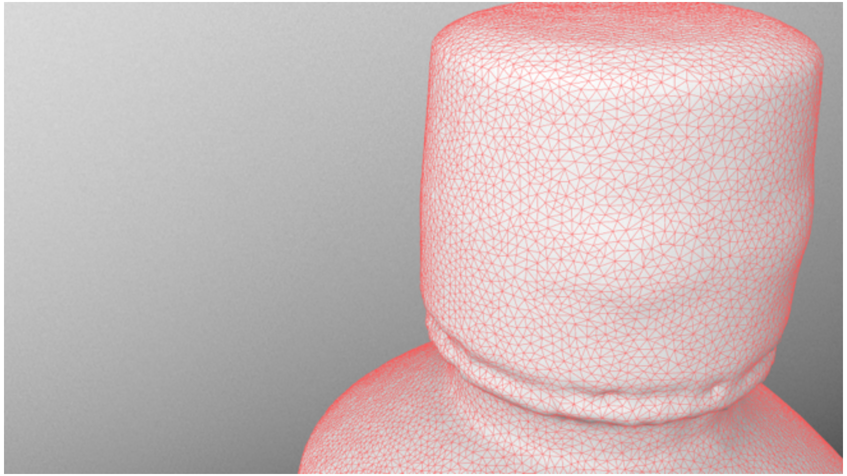
Our Approach: Final Mesh Creation



Kobbelt. CGIT 2000.
Lindstrom et. al. Visualization 1998.

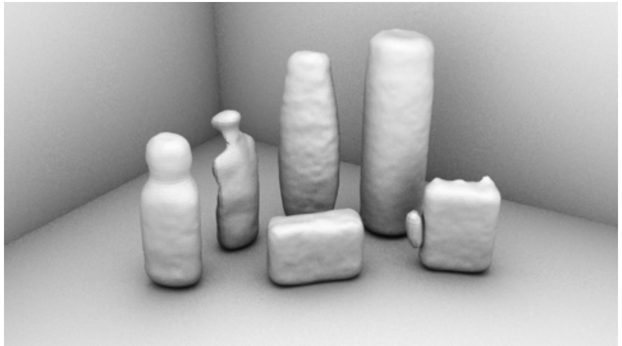


Before decifining



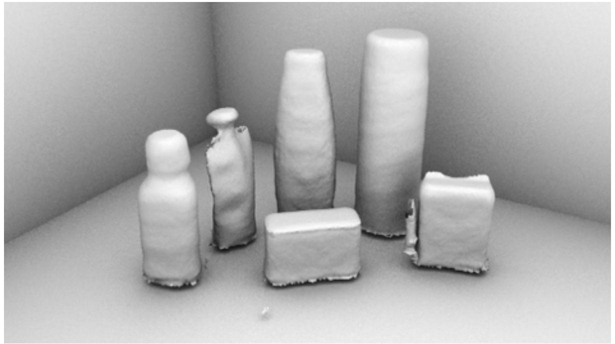
After decifining

Results (Simple Objects)

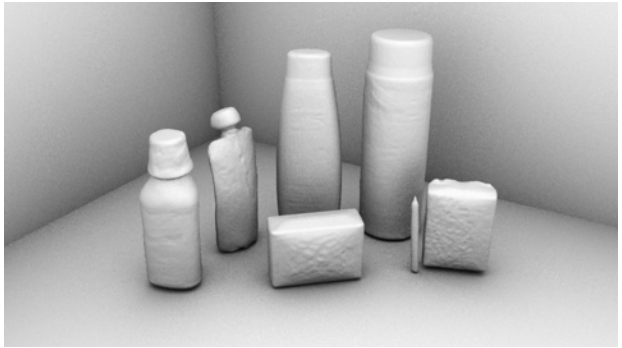


(a) PR

(b) VH

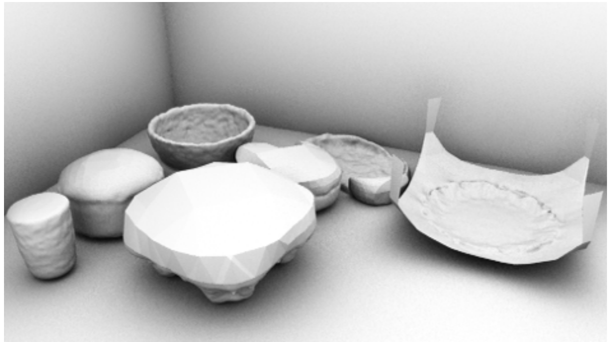


(c) KF



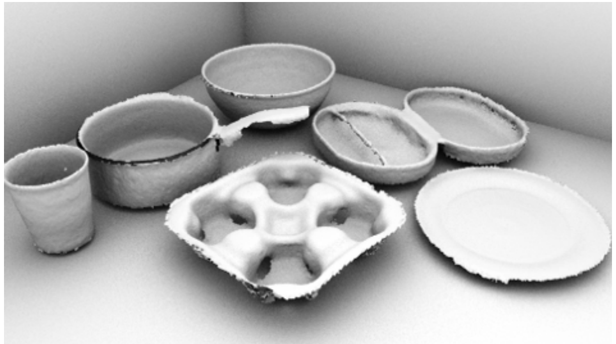
(d) Ours

Results (Objects with Concavities)

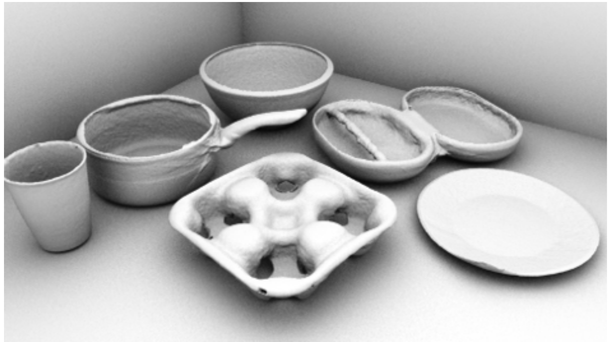


(a) PR

(b) VH

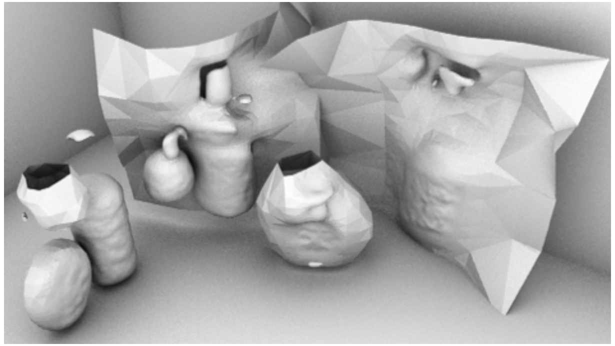


(c) KF



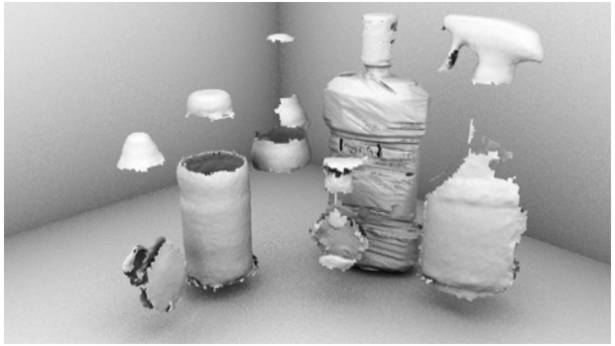
(d) Ours

Results (Objects with Translucencies)



(a) PR

(b) VH



(c) KF

(d) Ours

Quantitative Measurements

	Primitive Fitting RMS Errors (mm)			
	PR [25]	SVH	KF [22]	OUR METHOD
Pringles	0.566	**0.541**	0.850	0.563
Dove Soap	0.995	0.981	1.123	**0.948**
Almond Can	0.339	0.303	0.662	**0.294**
3M Spray	2.018	1.971	2.189	**1.958**

(a) Almonds Can

(b) Dove Soap Box

(c) Pringles Can

(d) 3M Spray

3D Scanning - Components

- Shape estimation

Newcombe et. al. ISMAR 2011

Narayan et. al. ICRA 2015

Zhou et. al. CVPR 2014

- Color estimation

Hernandez, Thesis 2004

Zhou et. al. ACM TOG 2014

Rusu, ICRA 2011

Previous Work: Comparisons

Original Image

PCL Volumetric Blending [1]

Saturation Blending [2]

Color Optimization [3]

[1] Rusu, ICRA 2011
[2] Hernandez et. al., Thesis 2004
[3] Zhou et. al., ACM TOG 2014

Previous Work: Comparisons

Original Image | **Our Approach** | **PCL Volumetric Blending [1]** | **Saturation Blending [2]** | **Color Optimization [3]**

[1] Rusu, ICRA 2011
[2] Hernandez et. al., Thesis 2004
[3] Zhou et. al., ACM TOG 2014

Our Approach

- We are given a mesh **M** which contains vertex set **P**.
- We are also given RGB images $\{I_i\}$ that observe the object.
- I_i has intrinsics matrix $\mathbf{K_i}$ and extrinsics matrix $\mathbf{T_i}$.
- For each $\mathbf{p} \in \mathbf{P}$, we want to estimate $\mathbf{C}(\mathbf{p})$, the color of vertex **p**. **C** denotes the color model.
- **p** is denotes a vertex position while $\mathbf{C}(\mathbf{p})$ is a 3-vector that represents an RGB value.

Our Approach

- $V(\mathbf{p}) \subset \{I_i\}$: the subset of images observing $\mathbf{p}$ without occlusion.
- $\mathbf{\Gamma_i}(\mathbf{p}, \mathbf{T_i})$: color obtained by projecting $\mathbf{p}$ onto I_i using extrinsics $\mathbf{T_i}$ and intrinsics $\mathbf{K_i}$.
- Error residual $|\mathbf{C}(\mathbf{p}) - \mathbf{\Gamma_i}(\mathbf{p}, \mathbf{T_i})|^2$ should be small.
- Viable objective to minimize:

$$\mathscr{J}(\mathbf{C},\mathbf{T}) = \frac{1}{2}\sum_{\mathbf{p}\in\mathbf{P}}\sum_{I_i\in V(\mathbf{p})} \|\vec{C}(\mathbf{p}) - \vec{\Gamma}_i(\mathbf{p},\mathbf{T}_i)\|^2$$

- Illumination assumptions?

Intermediate Results

- Current objective reduces ghosting
- Problems to resolve
  - Faded textures
  - Speckled regions

(a) Iteration 0 (b) Iteration 200

Intermediate Results

- Current objective reduces ghosting
- Problems to resolve
  - Faded textures
  - Speckled regions

(a) Iteration 0 (b) Iteration 200

Alleviating Faded Textures

- Original objective projects a vertex **v** onto *all* views
- Views with specularities will draw **v**'s color towards white, producing faded colors.
- Idea: select the top N views per vertex, where views are sorted by the camera foreshortening angle.

(a) Iteration 0 (b) Iteration 200

Alleviating Faded Textures

(a) N = 1 (b) N = 10 (c) N = 30 (c) N = 50 (d) N = 100

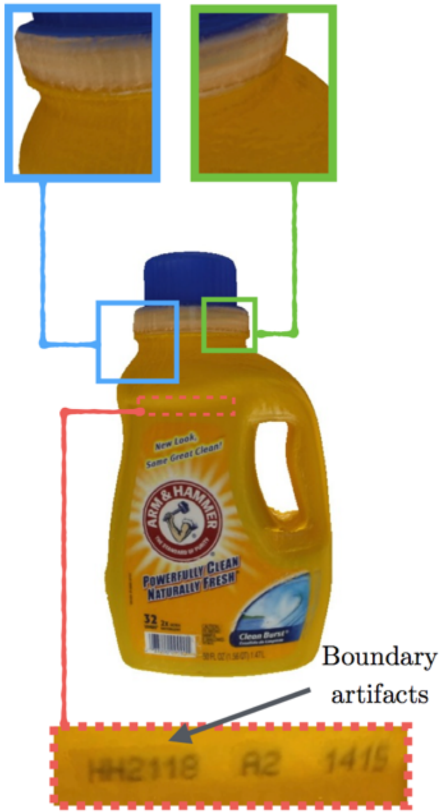
Intermediate Results

- Current objective reduces ghosting
- Problems to resolve
  - Faded textures
  - Speckled regions

(a) N = 1 (b) N = 10 (c) N = 30 (c) N = 50 (d) N = 100

Smoothing Speckled Regions

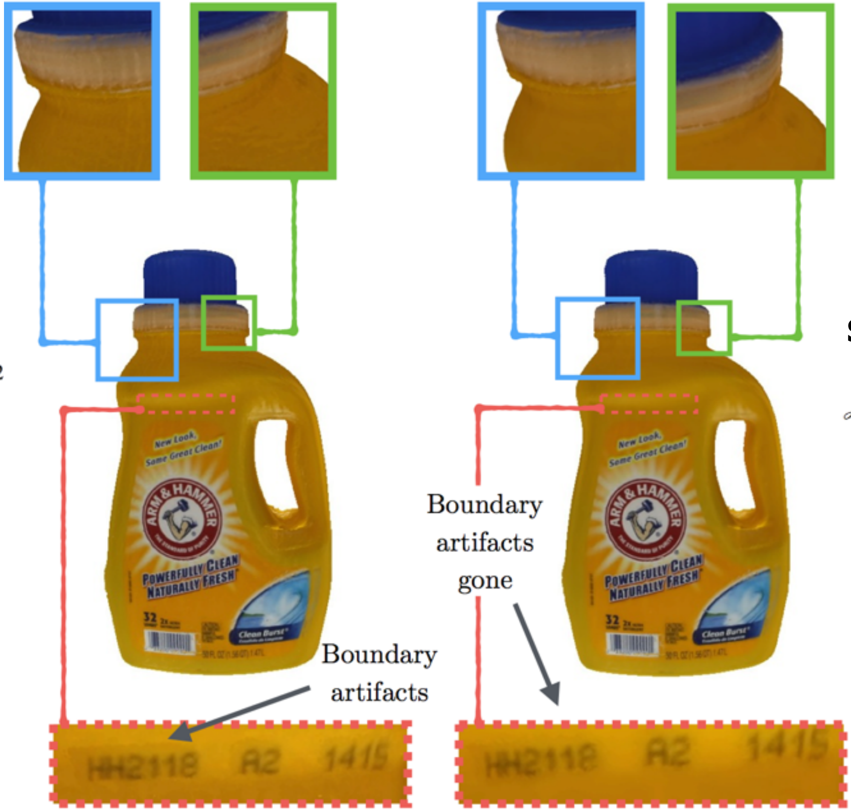
- Causes for speckled regions
  - Varying the cameras used in adjacent vertices
  - Differing N for adjacent vertices
- Idea: add color regularization term to the objective. We only smooth edges where:
  - Both vertices are not textured
  - Exactly one vertex is textured



Smoothing Speckled Regions

No smoothing

$$\mathcal{J}(\mathbf{C},\mathbf{T}) = \frac{1}{2}\sum_{\mathbf{p}\in\mathbf{P}}\sum_{I_i\in V'(\mathbf{p};t_\mathbf{p})} \|\vec{C}(\mathbf{p}) - \vec{\Gamma}_i(\mathbf{p},\mathbf{T}_i)\|^2$$



(a) No smoothing, λ = 0 (b) Smoothing, λ = 10

Smoothing

$$\mathcal{J}(\mathbf{C},\mathbf{T}) = \frac{1}{2}\sum_{\mathbf{p}\in\mathbf{P}}\sum_{I_i\in V'(\mathbf{p};t_\mathbf{p})} \|\vec{C}(\mathbf{p}) - \vec{\Gamma}_i(\mathbf{p},\mathbf{T}_i)\|^2 + \frac{\lambda}{2}\sum_{\mathbf{p}\in\mathbf{P}}\sum_{\mathbf{p}'\in N(\mathbf{p})} (1 - t_\mathbf{p}t_{\mathbf{p}'}) \cdot \|\vec{C}(\mathbf{p}) - \vec{C}(\mathbf{p}')\|^2$$

Alternating Optimization

$$\mathcal{J}(\mathbf{C},\mathbf{T}) = \frac{1}{2}\sum_{\mathbf{p}\in\mathbf{P}}\sum_{I_i\in V'(\mathbf{p};t_{\mathbf{p}})} \|\vec{C}(\mathbf{p}) - \vec{\Gamma}_i(\mathbf{p},\mathbf{T}_i)\|^2 + \frac{\lambda}{2}\sum_{\mathbf{p}\in\mathbf{P}}\sum_{\mathbf{p}'\in N(\mathbf{p})} (1 - t_{\mathbf{p}} t_{\mathbf{p}'}) \cdot \|\vec{C}(\mathbf{p}) - \vec{C}(\mathbf{p}')\|^2$$

- Idea: inspired from Zhou et. al. 2014, optimize **C** and **T** separately.

- Fixing **T** and optimizing **C**
  - Involves solving a quadratic objective.
- Fixing **C** and optimizing **T**
  - Gauss-Newton: involves solving ||I|| 6 x 6 systems of equations in parallel.

Coarse-to-fine Levenberg-Marquardt

- Resolving small features requires us to subdivide **M** multiple times.
- Sufficient subdivision yields meshes with 10x or more vertices.
- This substantially slows down optimization.
- Idea: run optimization in coarse-to-fine steps.

Original

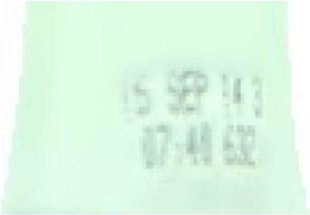
Reconstructed

Reconstructed, Zoom

Successes


Coca-Cola
REFRESCO
CONT. NET. 355 ml

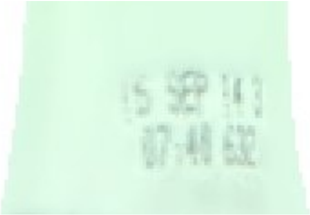

Coca-Cola
REFRESCO
CONT. NET. 355 ml


15 SEP 14 3
07:40 632

Successes


Coca-Cola
REFRESCO
CONT. NET. 355 ml


Coca-Cola
REFRESCO
CONT. NET. 355 ml




V8
V-FUSION
Vegetable & Fruit
Peach
Mango
100% JUICE


V8
V-FUSION
Peach
Mango
100% JUICE


Butternut Squash
& Pear
baby food


Butternut Squash
& Pear
baby food

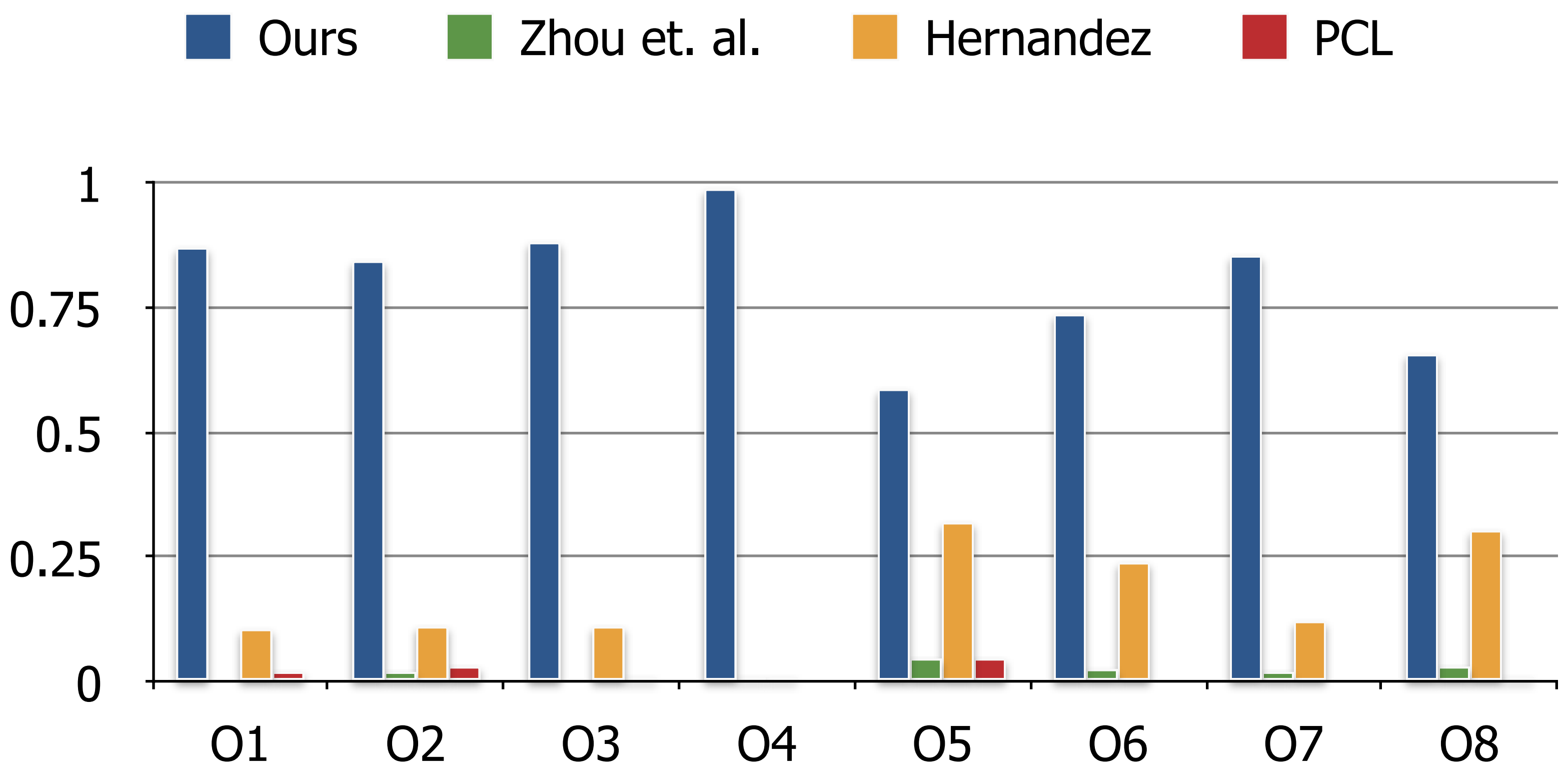
Cases to Improve

Cases to Improve

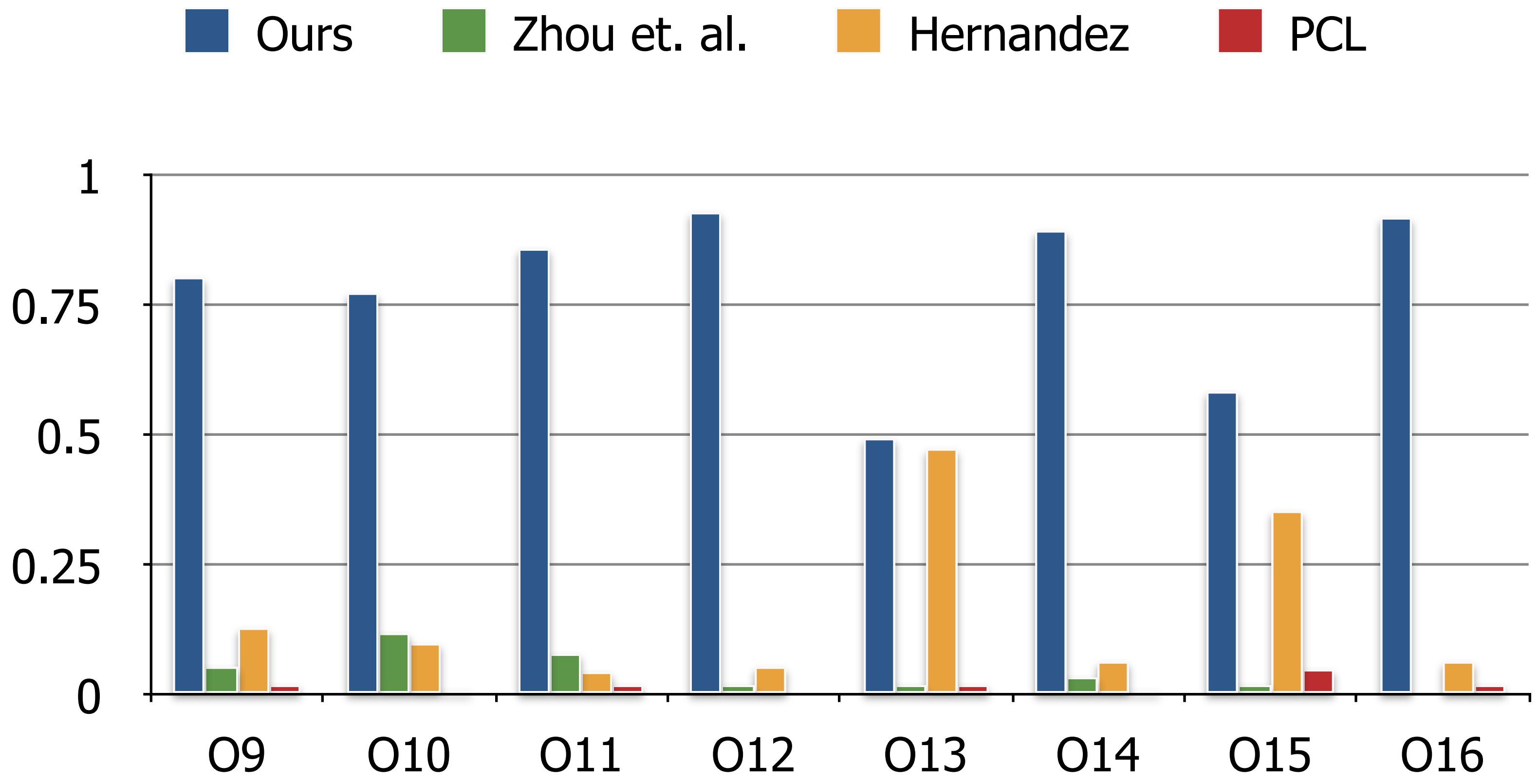
Evaluation Methodology

- We quantitatively compare our method to other techniques via an online user survey
  - http://tinyurl.com/iros2015coloropt
- Each participant is given 16 multiple choice questions
- Each question features a different reconstructed object
- Question asked: which of the following images matches "Reference" most closely?

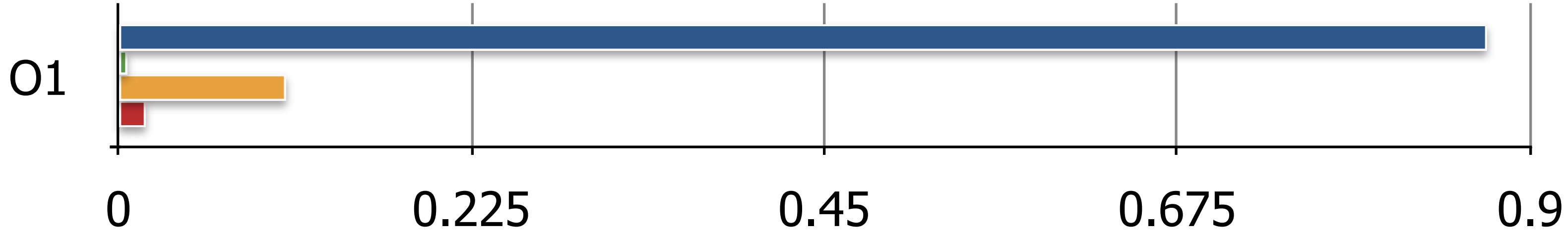
User Study Summary



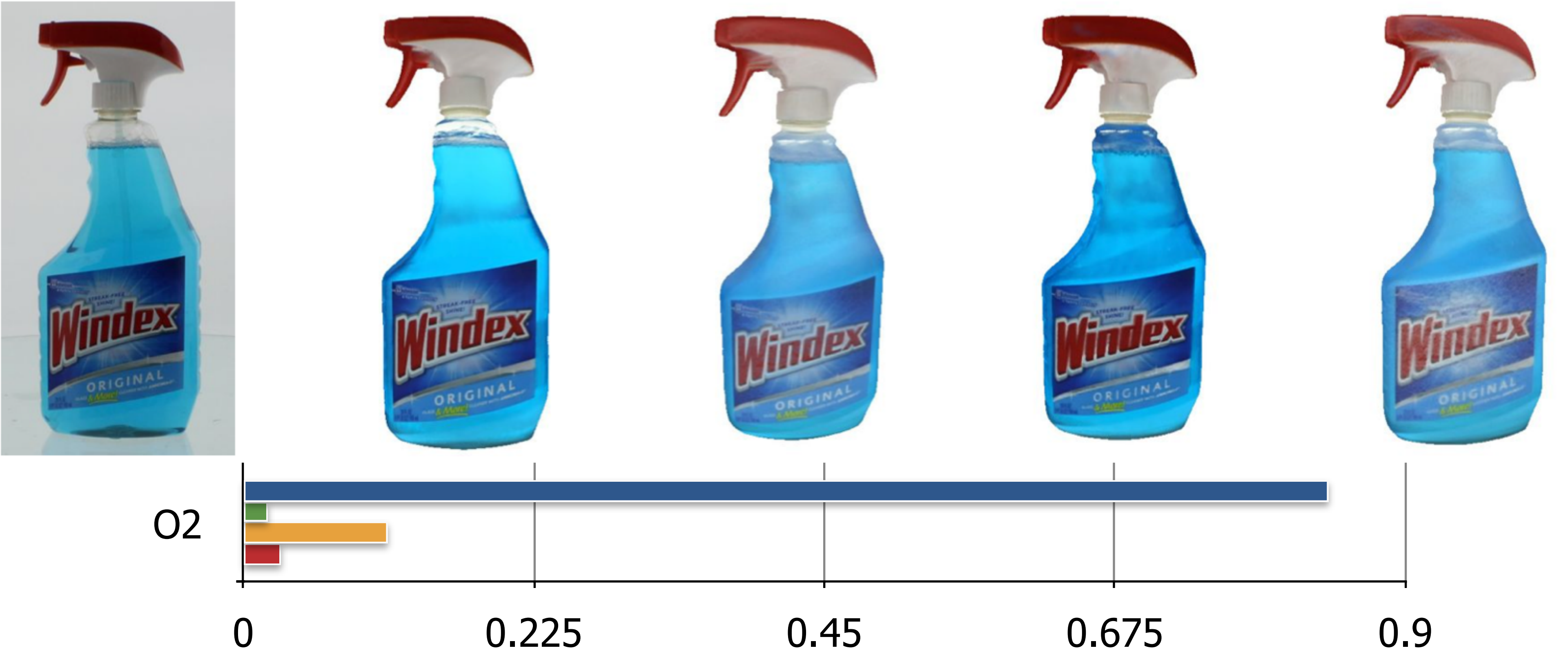
User Study Summary



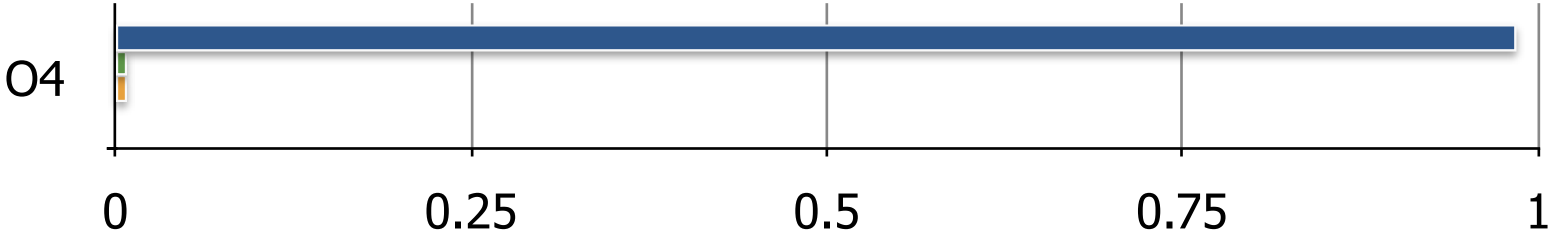
User Study Summary: O1



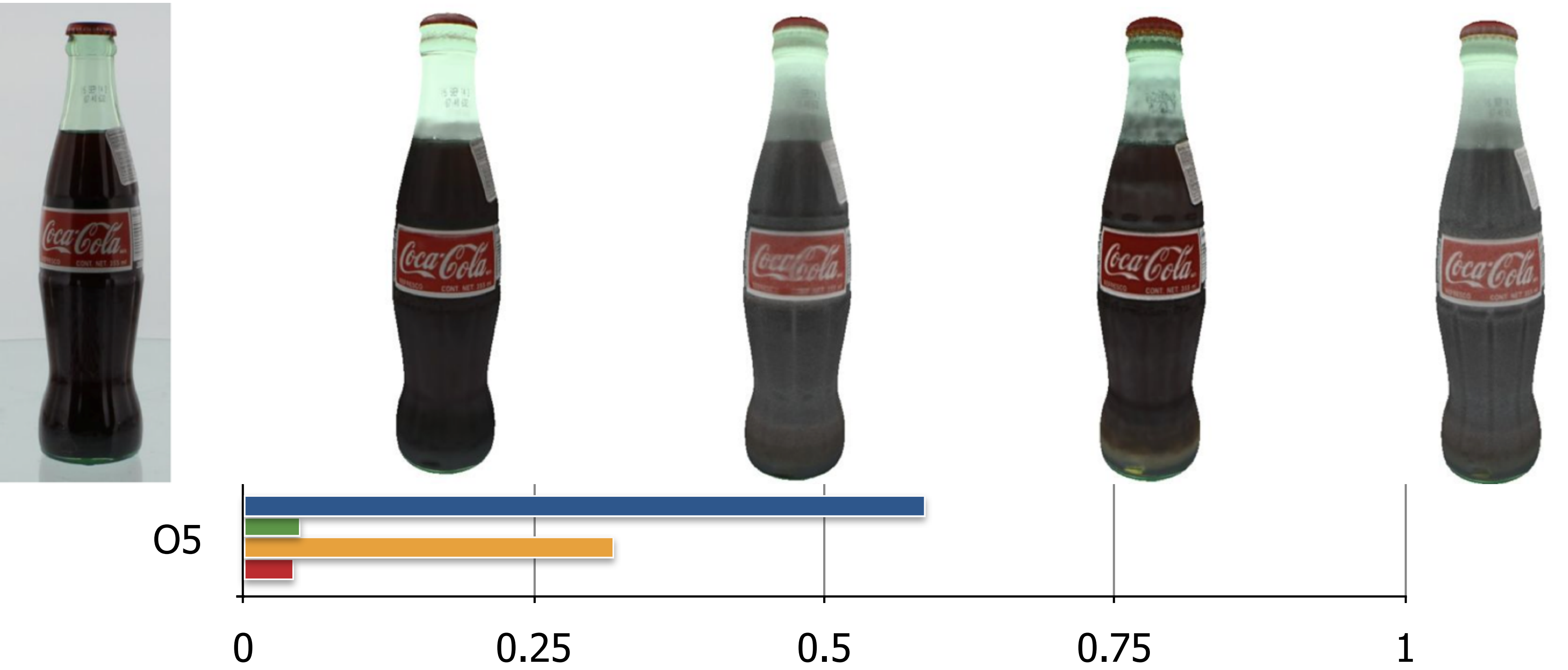
User Study Summary: O2



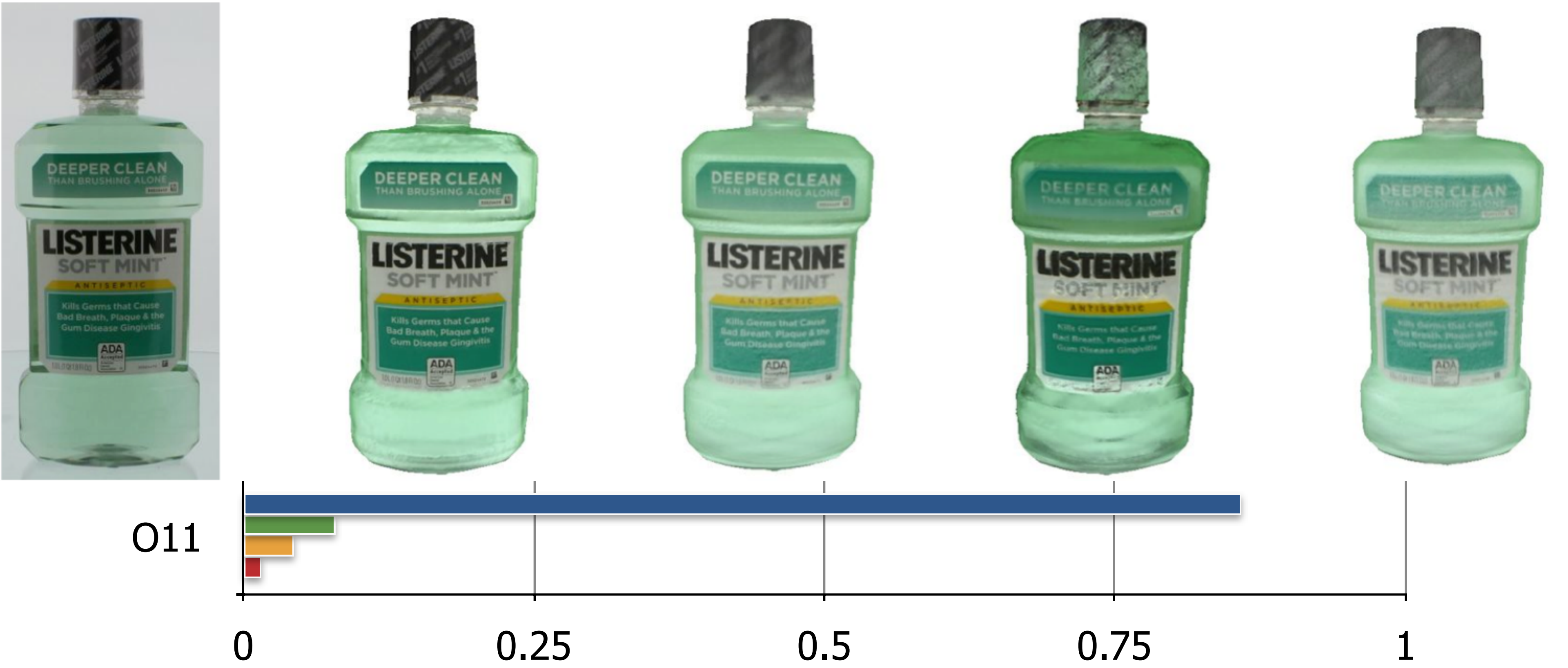
User Study Summary: O4



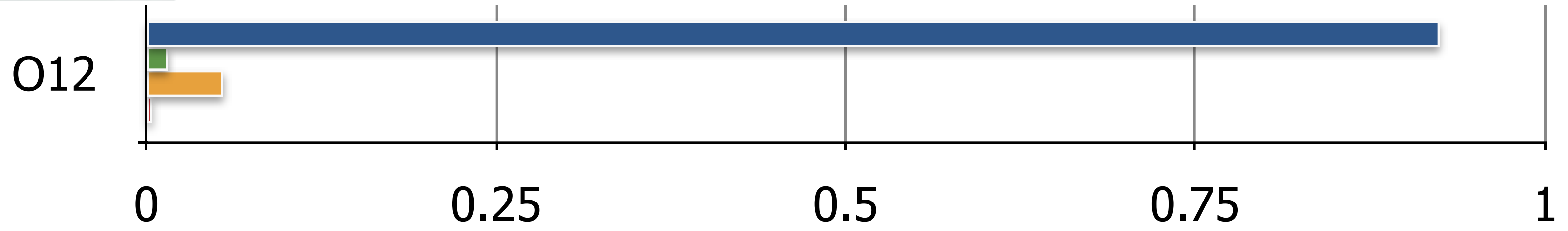
User Study Summary: O5



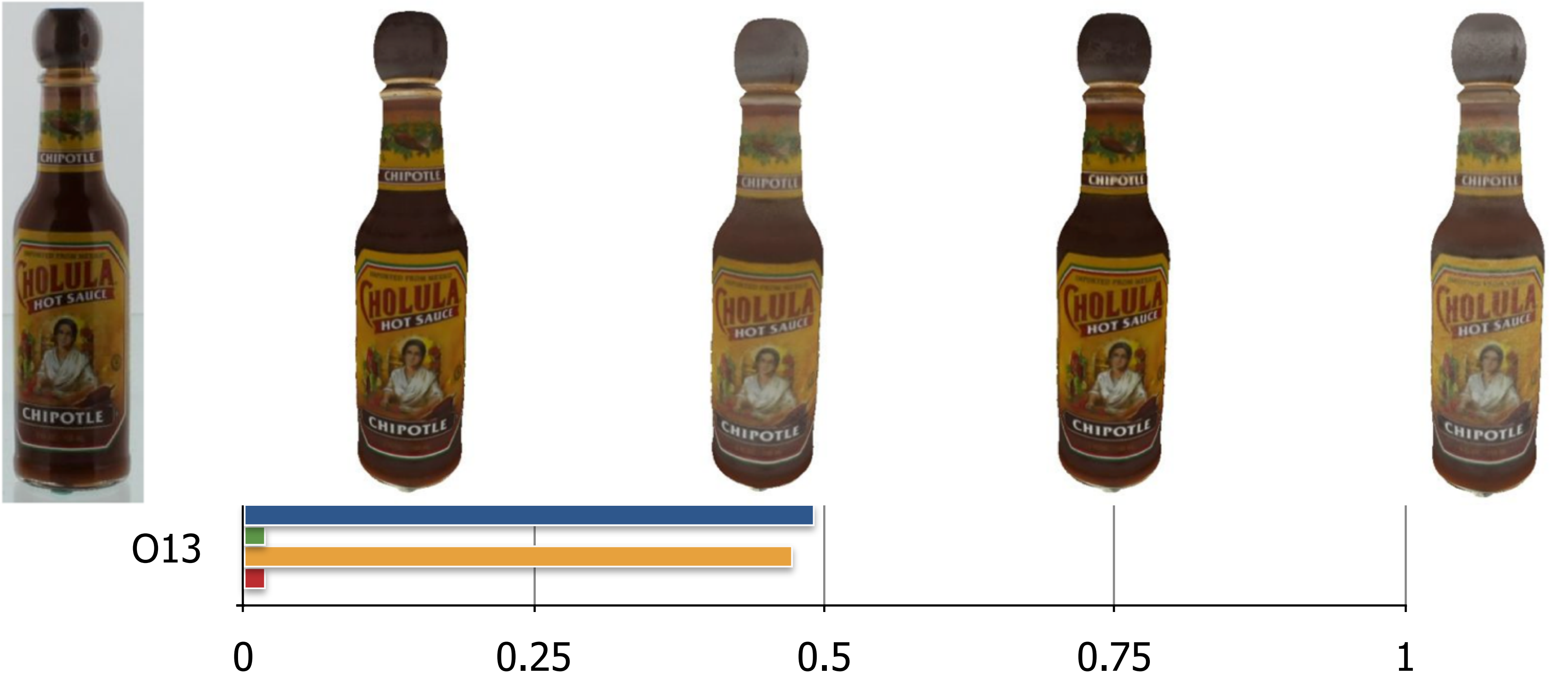
User Study Summary: O11



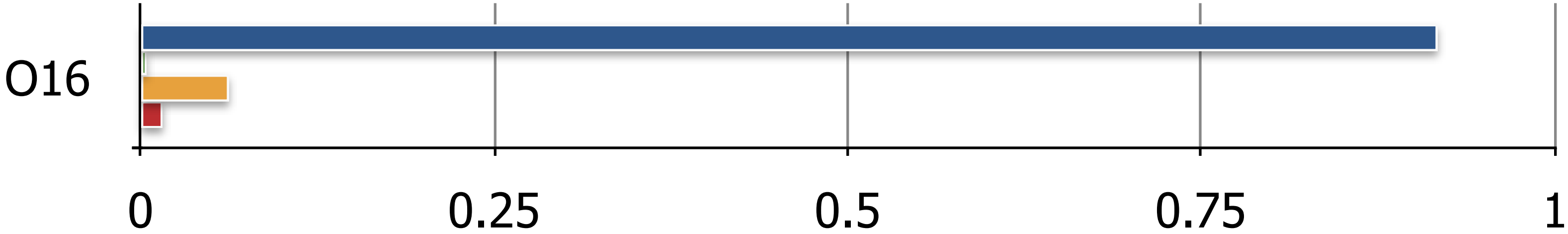
User Study Summary: O12



User Study Summary: O13

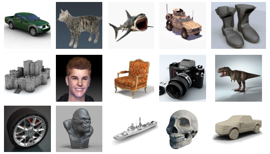


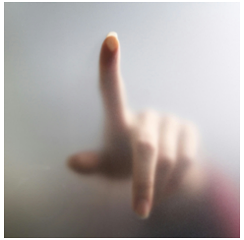
User Study Summary: O16



Scalable High-quality 3D Scanning

- Scalability
  - How do we scan many objects?
  - How do we keep device costs low?

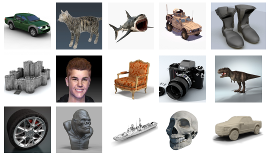


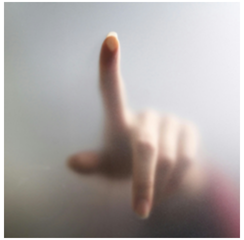


- High quality
  - How do we extract high quality shape models?
  - How do we extract high quality color models?

Scalable High-quality 3D Scanning

- Scalability
  - How do we scan many objects?
  - How do we keep device costs low?
- High quality
  - How do we extract high quality shape models?
  - How do we extract high quality color models?





Possible Future Directions

- Single-camera scanning with previous algorithms?
- Pitfalls:
  - Calibration accuracy
  - Loop closure issues
  - Manual intervention

Thank You