The First SUMO Challenge Workshop

ACCV 2018

Daniel Huber (Facebook), Lyne Tchampi (Stanford),
and Frank Dellaert (Georgia Tech)

SUMO Workshop Overview

Organizers

Daniel Huber
Facebook

Lyne Tchapmi
Stanford

Frank Dellaert
Georgia Tech

Focus and Objectives

- Scene Understanding and Modeling in 3D
- Discussions about the SUMO Challenge
- Current work from researchers in the field

SUMO Workshop Schedule

Speakers

Shuran Song
Columbia

Ian Reid
University of Adelaide

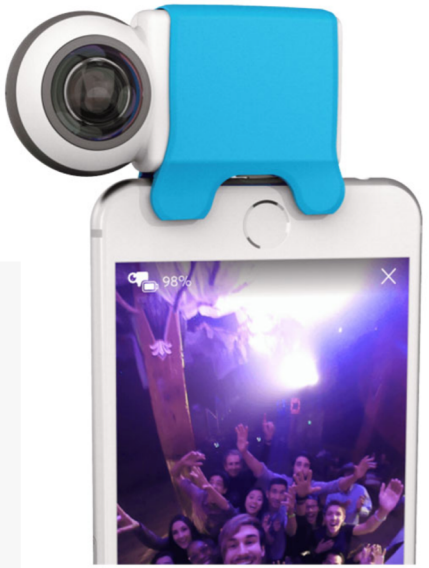
Rick Skarbez
La Trobe University

In So Kweon
KAIST

9:00 – 9:30 AM	Daniel Huber	Introduction and Overview of the SUMO Challenge
9:30 – 10:00 AM	Shuran Song	Comprehensive 3D Scene Understanding Beyond the Field of View
10:00 – 10:30	Break	
10:30 -- 11:00 AM	Ian Reid	SLAM in the Era of Deep Learning
11:00 – 11:30 AM	Rick Skarbez	Perception in SUMO: Justification and Experimental Results
11:30 – 12:00 PM	In So Kweon	Robust Image-based Modeling for Real World Applications

The Emergence of 360 and 180/360 RGBD Cameras

Giroptic IO



Insta 360 Pro

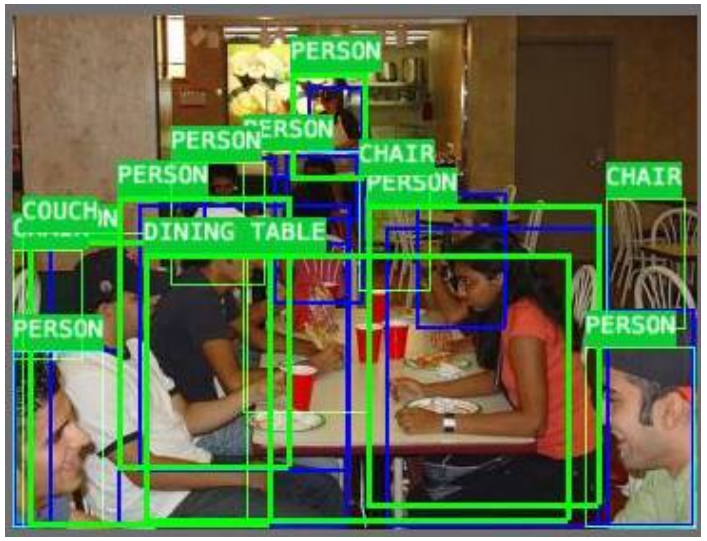
LucidCam

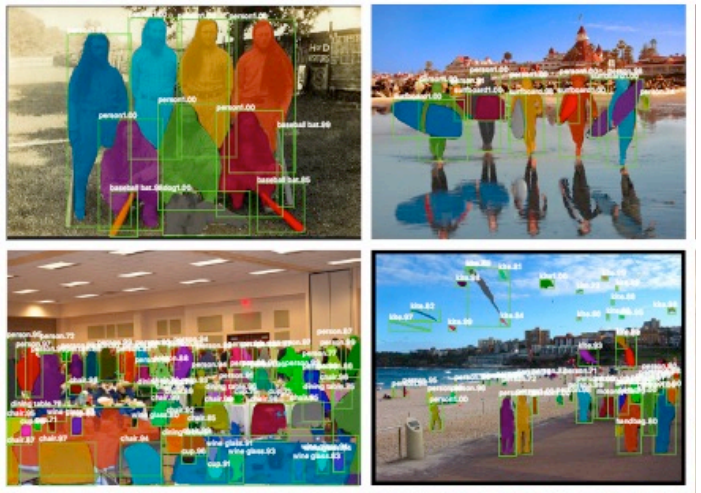
Ricoh Theta

Kandao

We Need to Redefine Scene Understanding

Today: Images





[image from HE ICCV 2017]

Tomorrow: 3D

[image from Song, SUNCG web site]

[image from Izadinia, CVPR 2016]

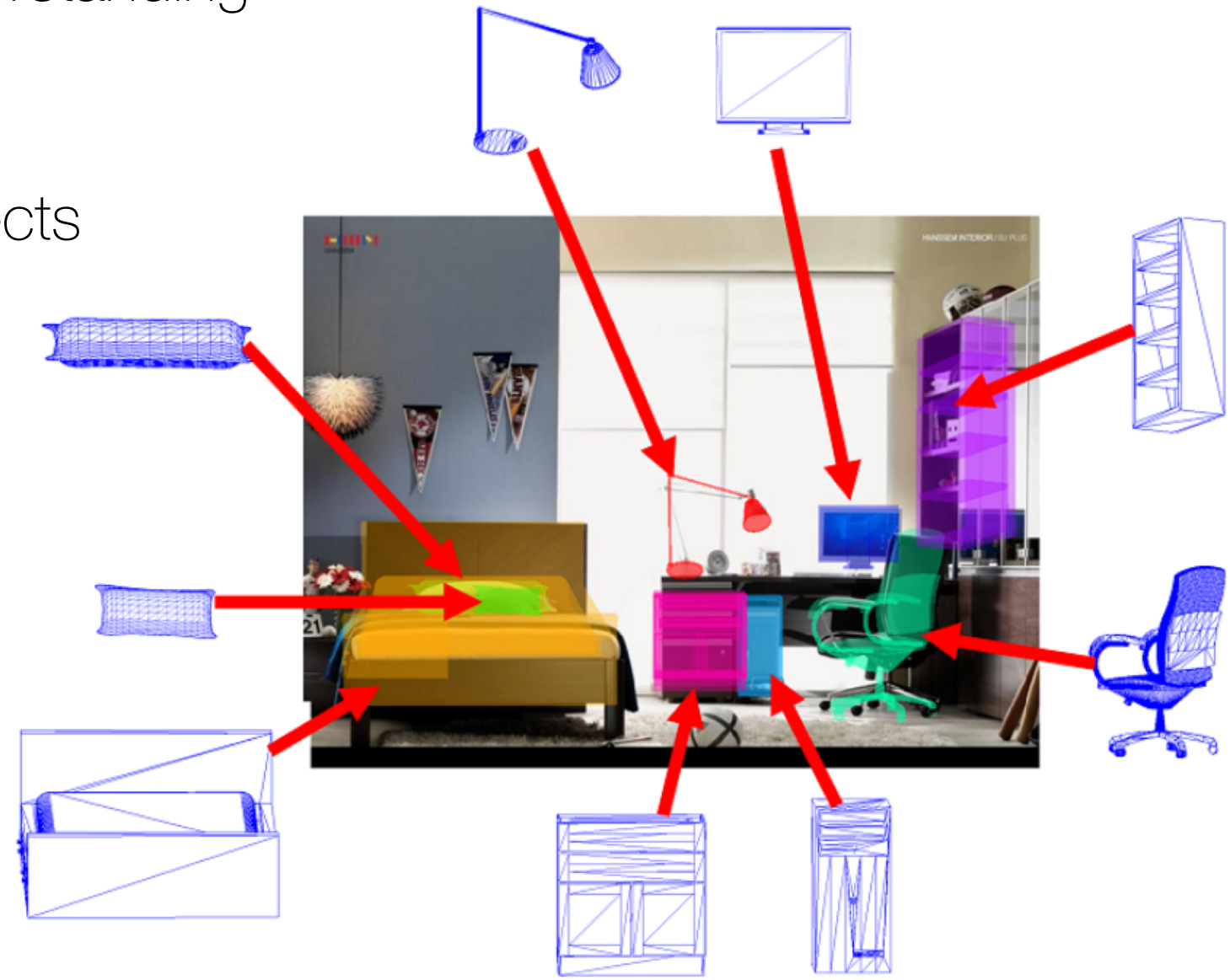
Object World

Comprehensive Scene Understanding

The world is made of objects

Objects have:

- Shape
- Pose
- Appearance
- Semantics



The SUMO Challenge

The SUMO Challenge

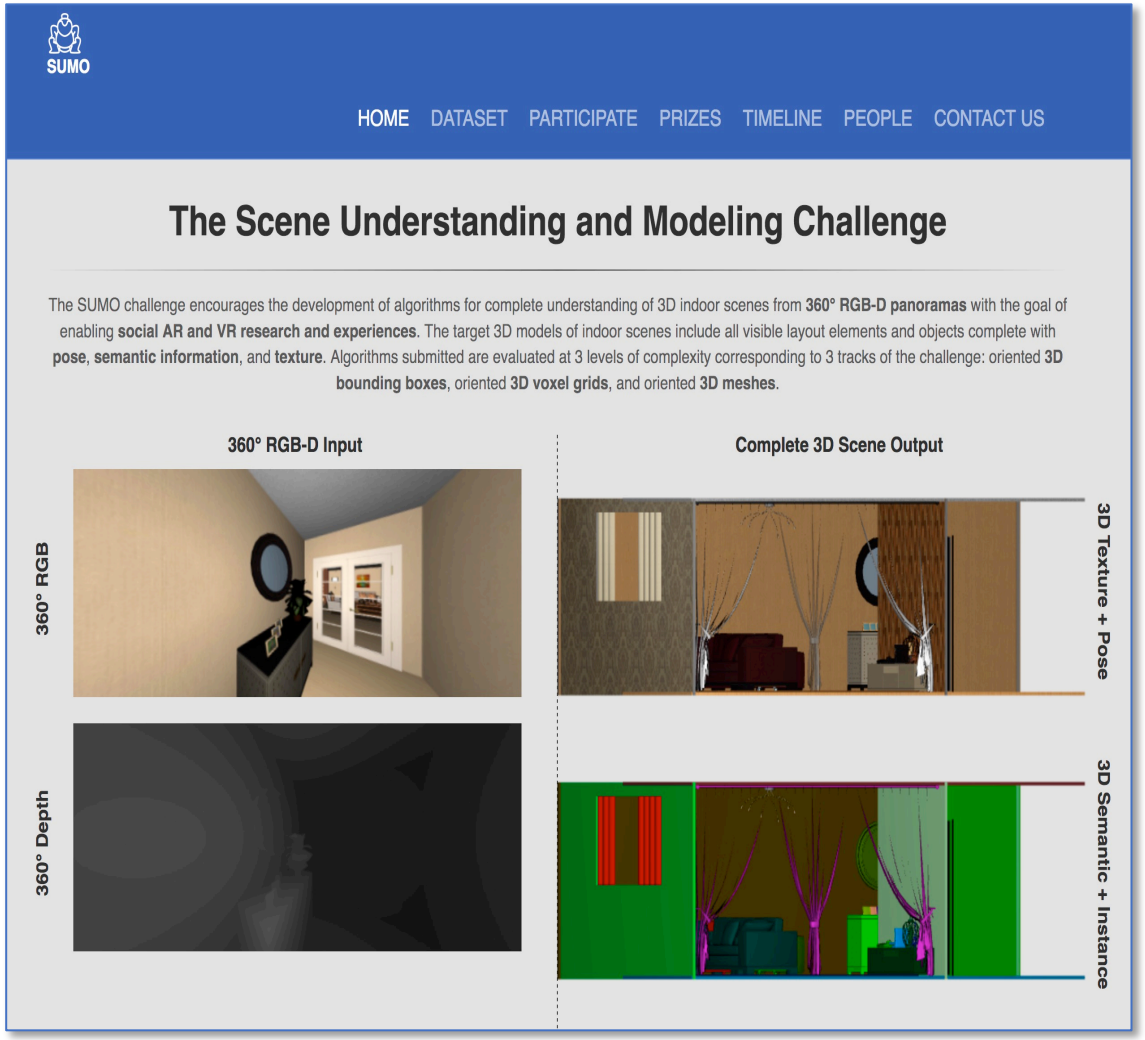
Comprehensive Scene Understanding from RGBD 360 Imagery

- Step 1 – Get a cool logo ✅

SUMO

THE SCENE UNDERSTANDING & MODELING CHALLENGE

- Step 2 – Get everyone to work on your problem



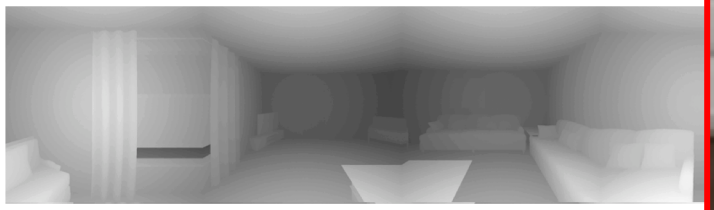
https://sumochallenge.org

SUMO Input and Output

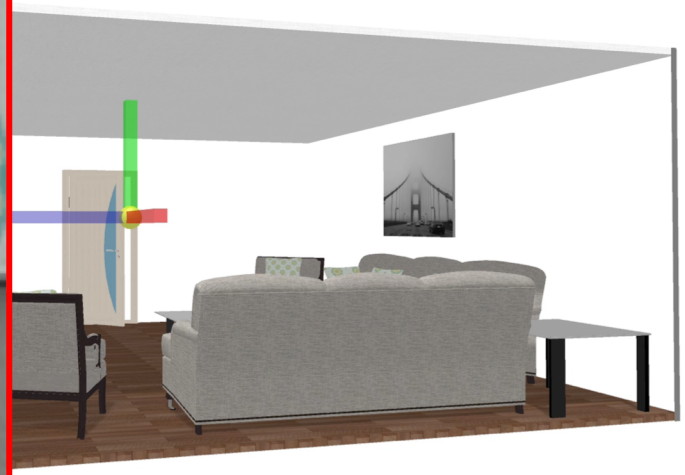
Input

Color

Depth



Output



SUMO Performance Tracks

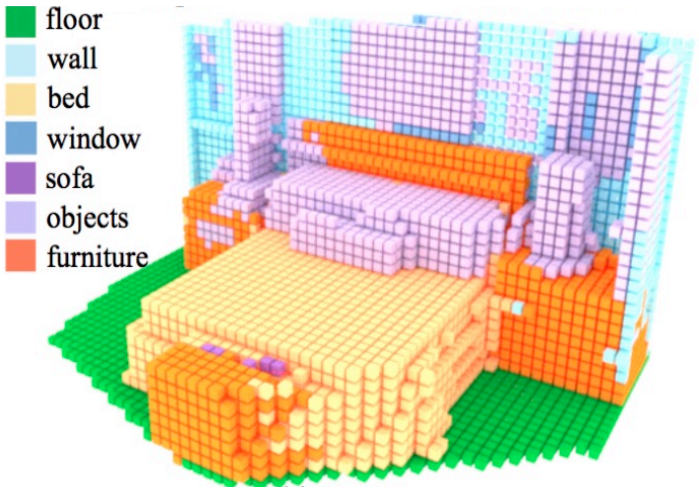
3D Bounding Box Track

- 3D bounding box
- 3D object pose
- Semantic category of element

[image from Song, CVPR 2015]

3D Voxel Grid Track

- 3D bounding box
- 3D object pose
- Semantic category of element
- Location and RGB color of occupied 3D voxels



[image from Song, CVPR 2017]

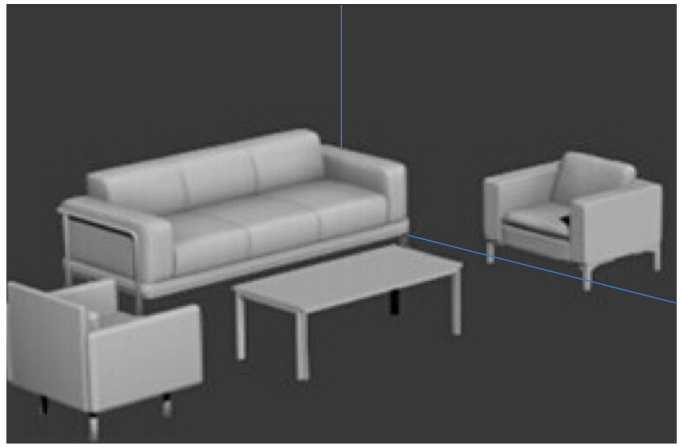
3D Mesh Track

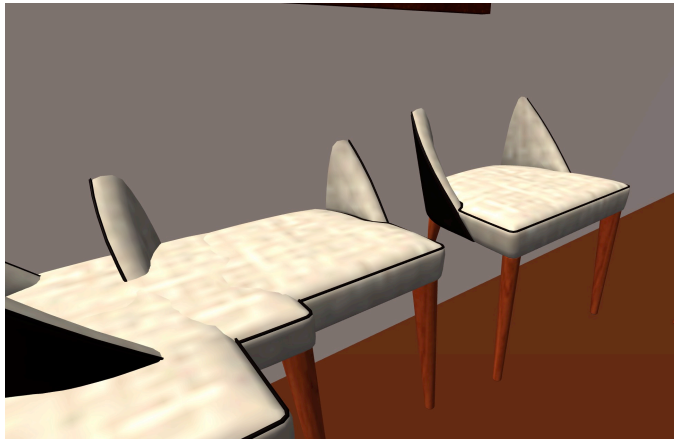
- 3D bounding box
- 3D object pose
- Semantic category of element
- Element's textured mesh (glb format)

[image from Izadinia, CVPR 2016]

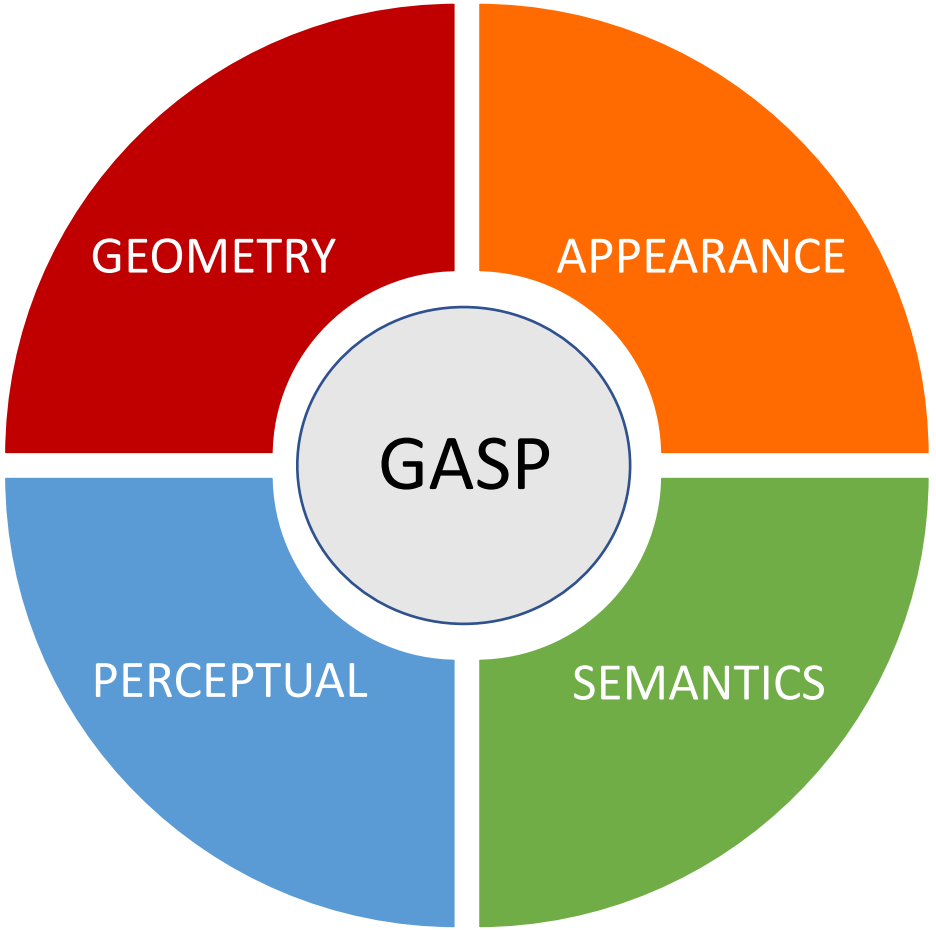
Evaluation – GASP Methodology

Shape and Pose





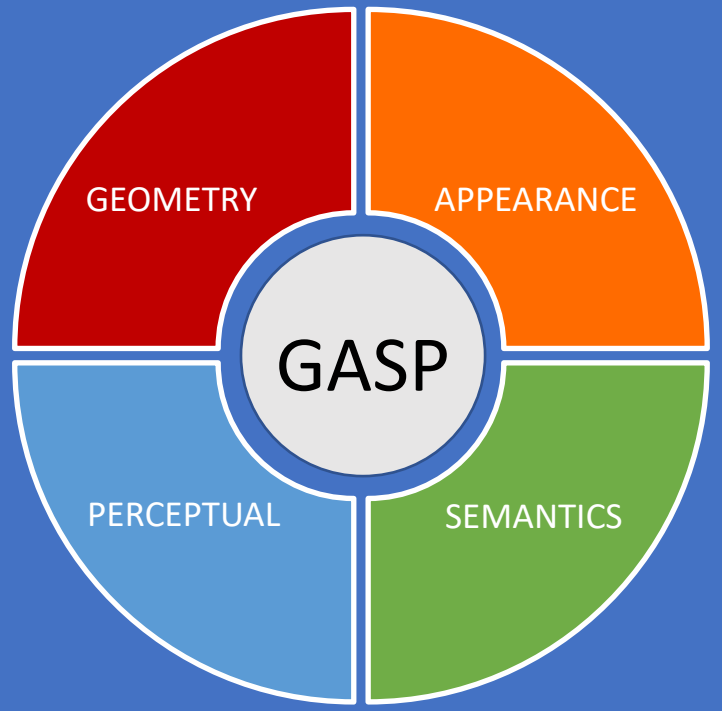
Weighted Perceptual Properties



Color

Mean Average Precision

Evaluation Metrics in Detail



Metrics Overview

		Track		
Metric		Bounding Boxes	Voxels	Meshes
Data Association				
Geometry	Shape			
	Pose			
Appearance				
Semantics				
Perceptual				

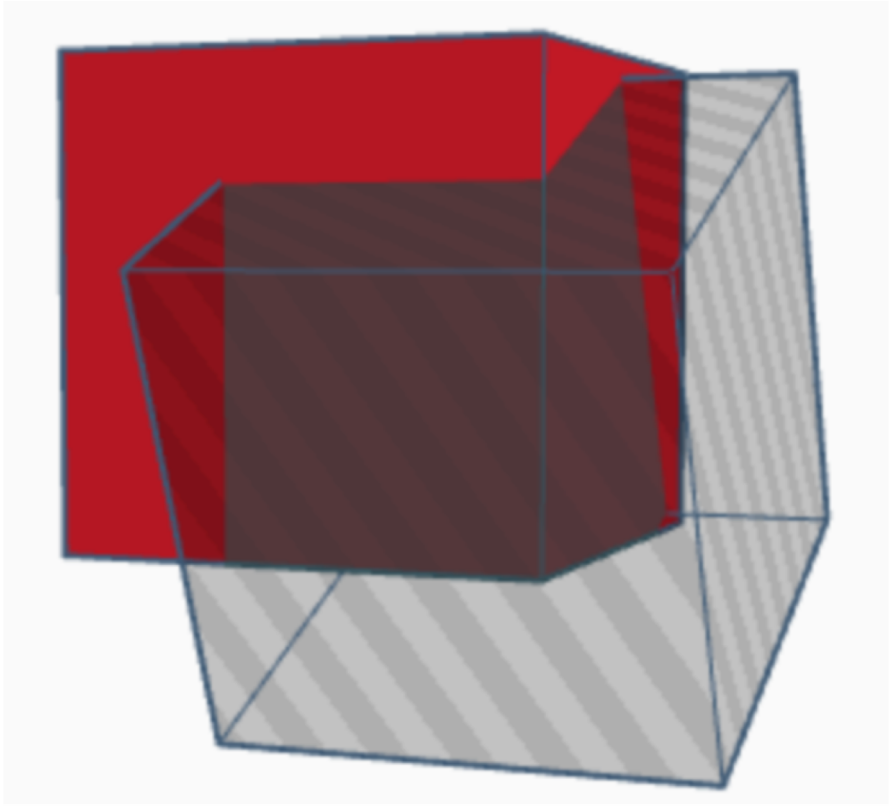
Data Association

		Track		
Metric		Bounding Boxes	Voxels	Meshes
Data Association		Greedy shape similarity		
Geometry	Shape			
	Pose			
Appearance				
Semantics				
Perceptual				

Data Association

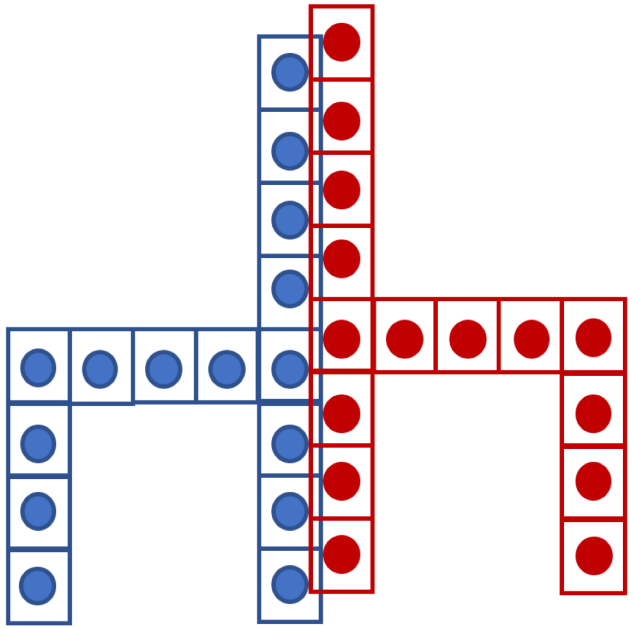
- Threshold on shape similarity
- Sort by detection score
- Greedy data association
  - 1:1 matching
  - Optionally constrained to match category

Shape Similarity – Bounding Box Track

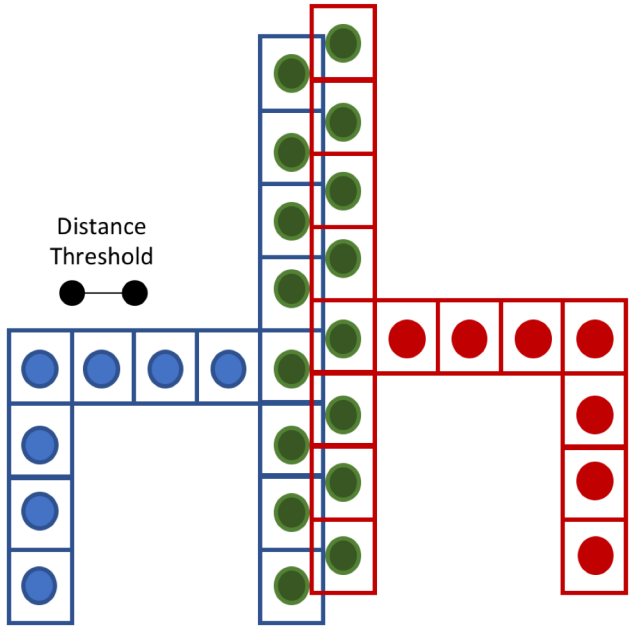


$$S_{\mathrm{bb}}(m,n) = \frac{V_{\mathrm{ov}_{mn}}}{V_m + V_n - V_{\mathrm{ov}_{mn}}}$$

Shape Similarity – Voxels Track



Voxelized Elements

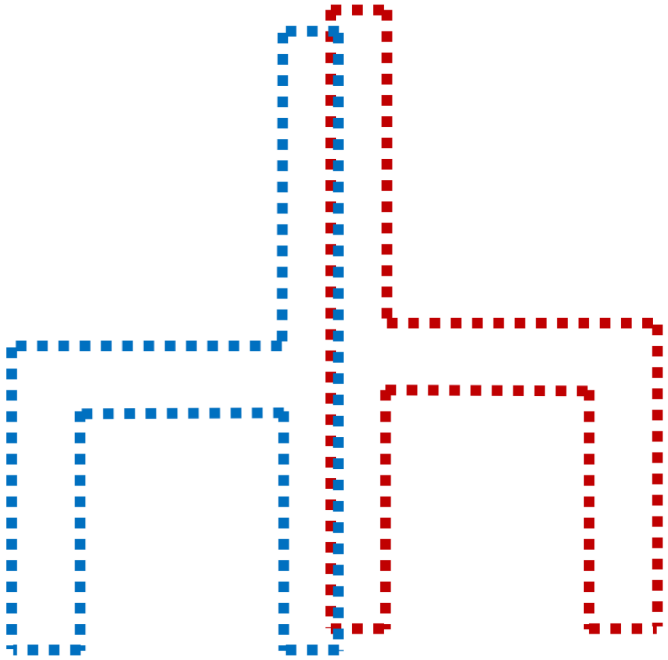


Intersection

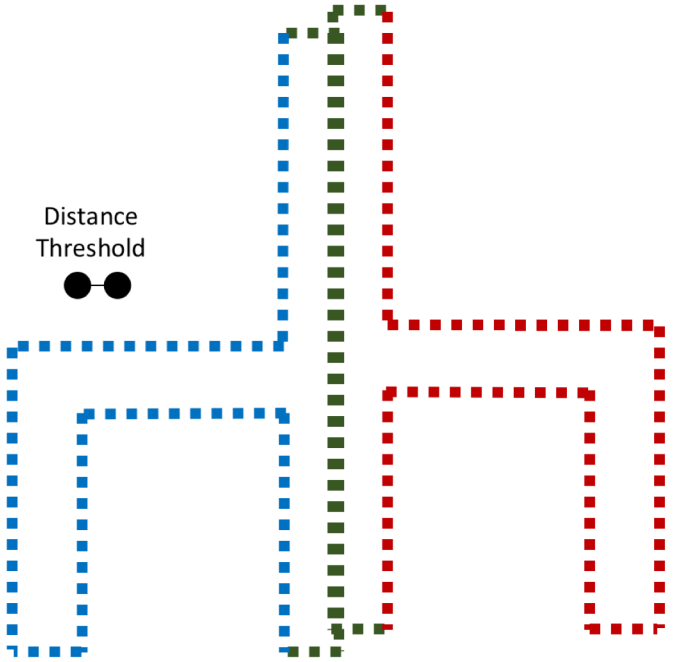
Union

$$S_{\text{vox}}(m,n) = \frac{|V_{c_{mn}}|}{|V_{c_m}| + |V_{c_n}|}$$

Shape Similarity – Meshes Track



Mesh Elements



Intersection

Union

$$S_{\text{mesh}}(m,n) = \frac{|B_{mn}|}{|B_m| + |B_n|}$$

Data Association

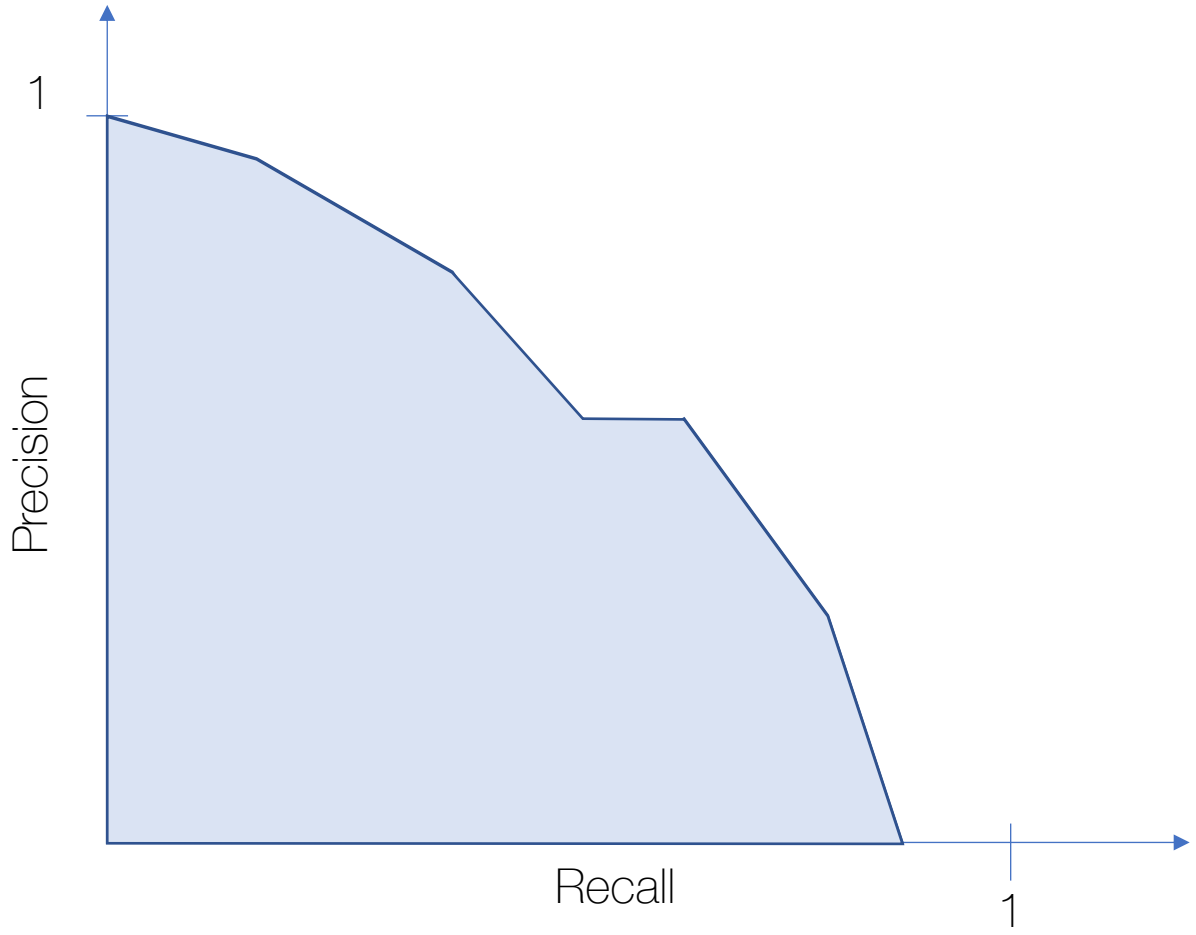
		Track		
Metric		Bounding Boxes	Voxels	Meshes
Data Association		Greedy shape similarity		
Geometry	Shape			
	Pose			
Appearance				
Semantics				
Perceptual				

Geometry Metrics – Shape

		Track		
Metric		Bounding Boxes	Voxels	Meshes
Data Association		Greedy shape similarity		
Geometry	Shape	Category-agnostic Mean Average Precision (mAP)		
	Pose			
Appearance				
Semantics				
Perceptual				

Geometry Metrics

Category-agnostic Mean Average Precision (mAP)



$$AP(\tau_i) = \frac{1}{11} \sum_{r \in \{0, 0.1, \ldots, 1\}} p_{\text{interp}}(r, \tau_i)$$

$$p_{\text{interp}}(r, \tau_i) = \max_{\tilde{r}: \tilde{r} \geq r} p(\tilde{r}, tau_i)$$

$$mAP = \frac{1}{|\tau|} \sum_{i=1}^{|\tau|} AP(\tau_i)$$

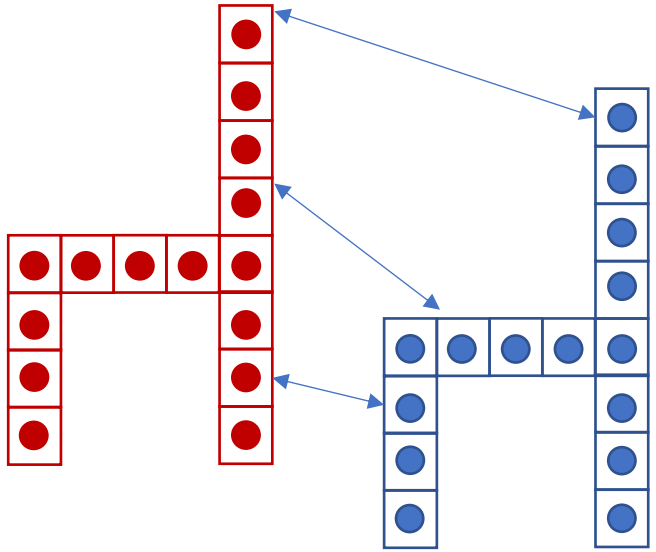
Geometry Metrics – Shape

		Track		
Metric		Bounding Boxes	Voxels	Meshes
Data Association		Greedy shape similarity		
Geometry	Shape	Category-agnostic Mean Average Precision (mAP)		
	Pose			
Appearance				
Semantics				
Perceptual				

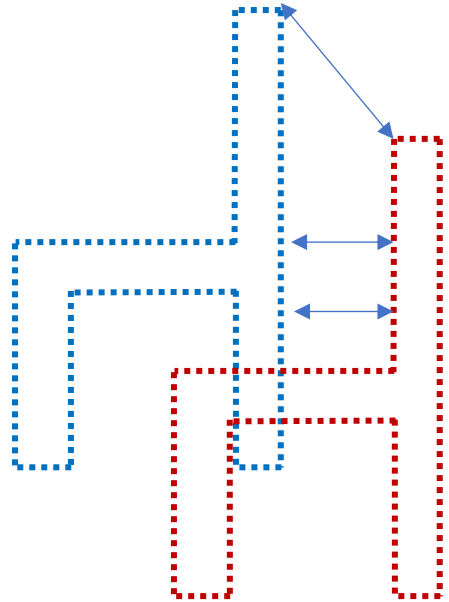
Geometry Metrics – Shape

		Track		
Metric		Bounding Boxes	Voxels	Meshes
Data Association		Greedy shape similarity		
Geometry	Shape	Category-agnostic Mean Average Precision (mAP)		
		N/A	Voxel RMSSSD	Surface point RMSSSD
	Pose			
Appearance				
Semantics				
Perceptual				

RMS Symmetric Shape Distance (RMSSSD)



Voxels



Meshes

$$RMSSSD = \frac{1}{|\tau|}\sum_{i=1}^{|\tau|}\frac{1}{|M_i|}\sum_{j=1}^{|M_i|}\frac{1}{\sqrt{|B_j|+|B_{jm_i}|}}\sqrt{\sum_{x\in B_j} d^2(x, B_{jm_i}) + \sum_{y\in B_{jm_i}} d^2(y, B_j)}$$

Geometry Metrics – Shape

		Track		
Metric		Bounding Boxes	Voxels	Meshes
Data Association		Greedy shape similarity		
Geometry	Shape	Category-agnostic Mean Average Precision (mAP)		
		N/A	Voxel RMSSSD	Surface point RMSSSD
	Pose			
Appearance				
Semantics				
Perceptual				

Geometry Metrics – Pose

		Track		
Metric		Bounding Boxes	Voxels	Meshes
Data Association		Greedy shape similarity		
Geometry	Shape	Category-agnostic Mean Average Precision (mAP)		
		N/A	Voxel RMSSSD	Surface point RMSSSD
	Pose	• Average geodesic distance (rotation) • Average translation error (translation)		
Appearance				
Semantics				
Perceptual				

Geometry Metrics – Pose

Rotation

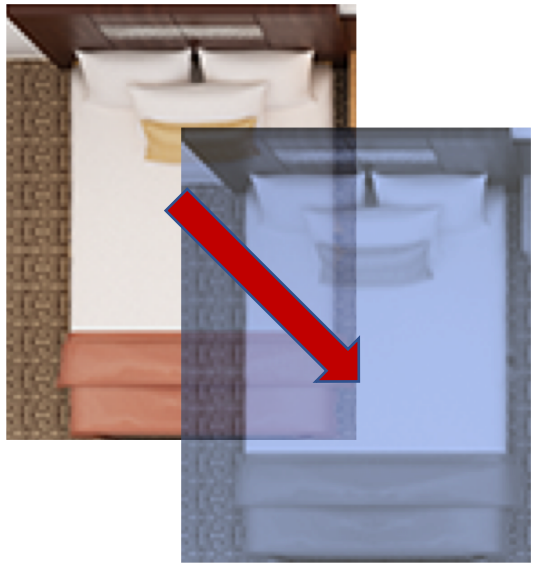
Average Geodesic Distance

$$\Delta R = \frac{1}{|\tau|}\sum_{i=1}^{|\tau|}\frac{1}{|M_i|}\sum_{j=1}^{|M_i|}\Delta r(R_{i,j}^{\mathrm{GT}}, R_{i,j}^{\mathrm{DET}})$$

$$\Delta r(R_a, R_b) \equiv \frac{1}{\sqrt{2}}||\log(R_a^T R_b)||_F$$

Translation

Average Translation Distance



$$\Delta T = \frac{1}{|\tau|}\sum_{i=1}^{|\tau|}\frac{1}{|M_i|}\sum_{j=1}^{|M_i|}\Delta t(t_{i,j}^{\mathrm{GT}}, t_{i,j}^{\mathrm{DET}})$$

$$\Delta t(t_a, t_b) = ||t_a - t_b||$$

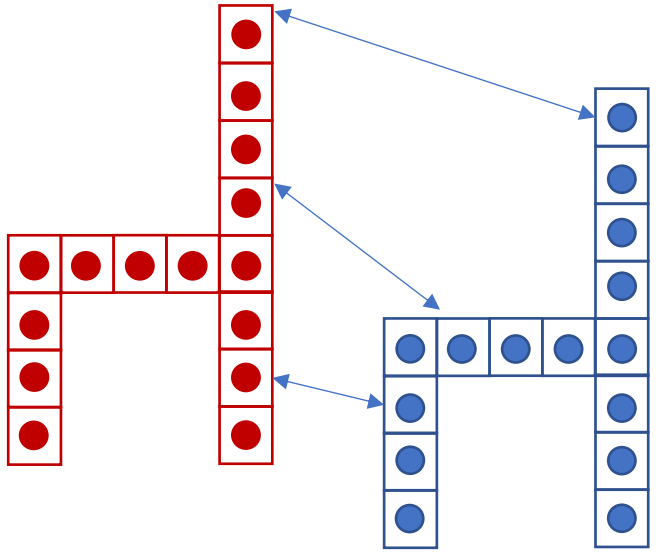
Geometry Metrics – Pose

		Track		
Metric		Bounding Boxes	Voxels	Meshes
Data Association		Greedy shape similarity		
Geometry	Shape	Category-agnostic Mean Average Precision (mAP)		
		N/A	Voxel RMSSSD	Surface point RMSSSD
	Pose	• Average geodesic distance (rotation) • Average translation error (translation)		
Appearance				
Semantics				
Perceptual				

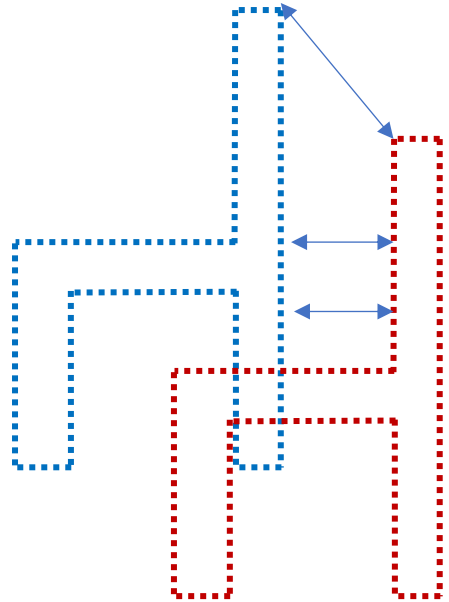
Appearance Metrics

		Track		
Metric		Bounding Boxes	Voxels	Meshes
Data Association		Greedy shape similarity		
Geometry	Shape	Category-agnostic Mean Average Precision (mAP)		
		N/A	Voxel RMSSSD	Surface point RMSSSD
	Pose	• Average geodesic distance (rotation) • Average translation error (translation)		
Appearance		N/A	RMS color distance (RMSSSCD)	
Semantics				
Perceptual				

RMS Symmetric Surface Color Distance (RMSSSCD)



Voxels



Meshes

$$RMSSSCD = \frac{1}{|\tau|}\sum_{i=1}^{|\tau|}\frac{1}{|M_i|}\sum_{j=1}^{|M_i|}\Delta d_{RGB}(j, jm_i)$$

$$\Delta d_{RGB}(j, jm_i) = \frac{1}{\sqrt{|B_j| + |B_{jm_i}|}}\sqrt{\sum_{x \in B_j} d^2_{RGB}(x, B_{jm_i}) + \sum_{y \in B_{jm_i}} d^2_{RGB}(y, B_j)}$$

Appearance Metrics

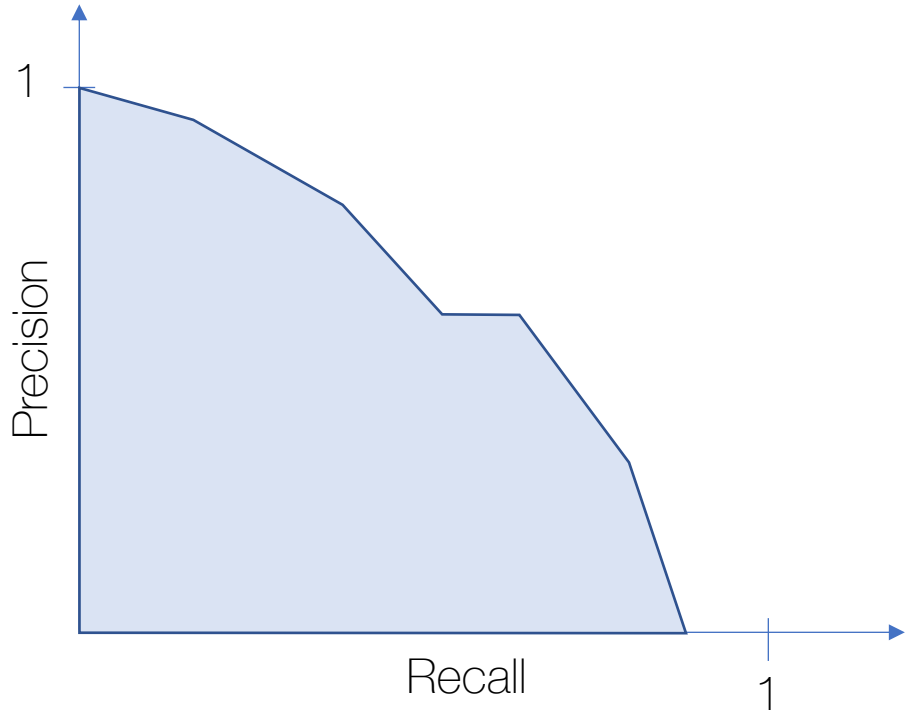
		Track		
Metric		Bounding Boxes	Voxels	Meshes
Data Association		Greedy shape similarity		
Geometry	Shape	Category-agnostic Mean Average Precision (mAP)		
		N/A	Voxel RMSSSD	Surface point RMSSSD
	Pose	• Average geodesic distance (rotation) • Average translation error (translation)		
Appearance		N/A	RMS color distance (RMSSSCD)	
Semantics				
Perceptual				

Semantic Metrics

		Track		
Metric		Bounding Boxes	Voxels	Meshes
Data Association		Greedy shape similarity		
Geometry	Shape	Category-agnostic Mean Average Precision (mAP)		
		N/A	Voxel RMSSSD	Surface point RMSSSD
	Pose	• Average geodesic distance (rotation) • Average translation error (translation)		
Appearance		N/A	RMS color distance (RMSSSCD)	
Semantics		Category-specific Mean Average Precision (mAP)		
Perceptual				

Semantic Metrics

Category-specific Mean Average Precision (mAP)



$$mAP = \frac{1}{|\tau|}\frac{1}{|C|}\sum_{i=1}^{|\tau|}\sum_{c\in C} AP(\tau_i, c)$$

$$AP(\tau_i, C_j) = \frac{1}{11}\sum_{r\in\{0,0.1,\ldots,1\}} p_{\text{interp}_{C_j}}(r, \tau_i)$$

$$p_{\text{interp}_{C_j}}(r, \tau_i) = \max_{\tilde{r}:\tilde{r}\geq r} p(\tilde{r}, tau_i)$$

Semantic Metrics

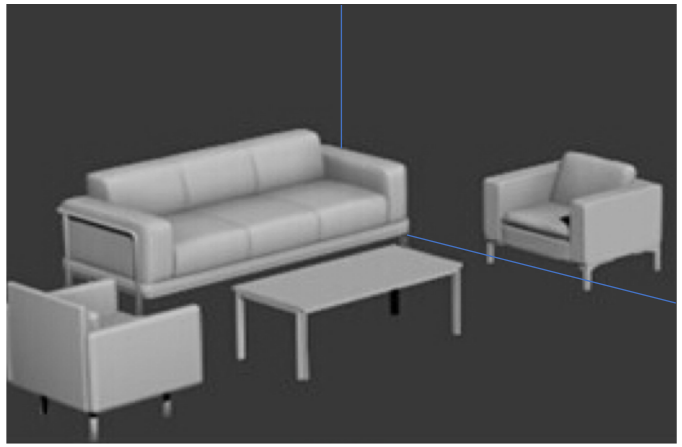
		Track		
Metric		Bounding Boxes	Voxels	Meshes
Data Association		Greedy shape similarity		
Geometry	Shape	Category-agnostic Mean Average Precision (mAP)		
		N/A	Voxel RMSSSD	Surface point RMSSSD
	Pose	• Average geodesic distance (rotation) • Average translation error (translation)		
Appearance		N/A	RMS color distance (RMSSSCD)	
Semantics		Category-specific Mean Average Precision (mAP)		
Perceptual				

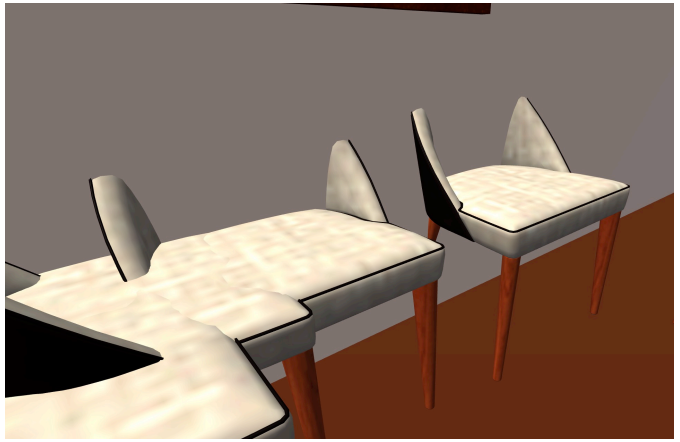
Perceptual Metrics

		Track		
Metric		Bounding Boxes	Voxels	Meshes
Data Association		Greedy shape similarity		
Geometry	Shape	Category-agnostic Mean Average Precision (mAP)		
		N/A	Voxel RMSSSD	Surface point RMSSSD
	Pose	• Average geodesic distance (rotation) • Average translation error (translation)		
Appearance		N/A	RMS color distance (RMSSSCD)	
Semantics		Category-specific Mean Average Precision (mAP)		
Perceptual		Average of weighted Gaussians		

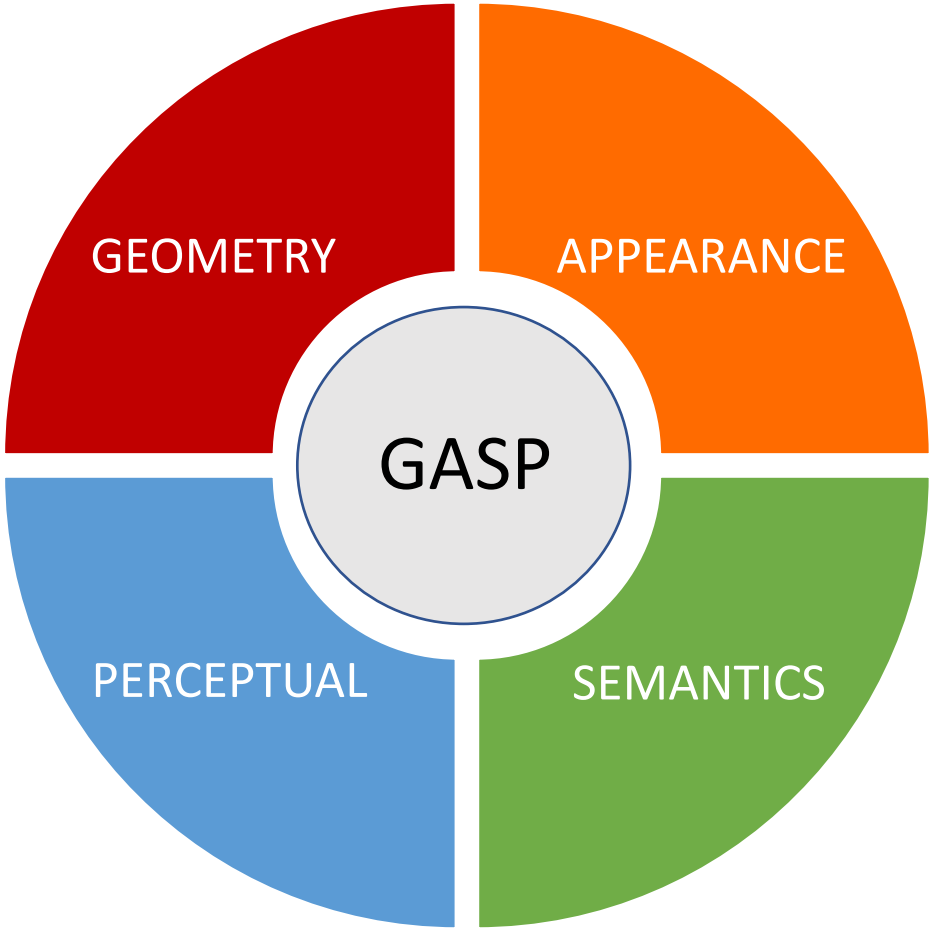
Evaluation – GASP Methodology

Shape and Pose





Weighted Perceptual Properties



Color

Mean Average Precision

SUMO Logistics

Timeline

Date	Event
June 22, 2018	Challenge announcement
July 30, 2018	Challenge launch / data release
December 3, 2018	SUMO Workshop @ ACCV 2018
December 21, 2018	Final submission deadline

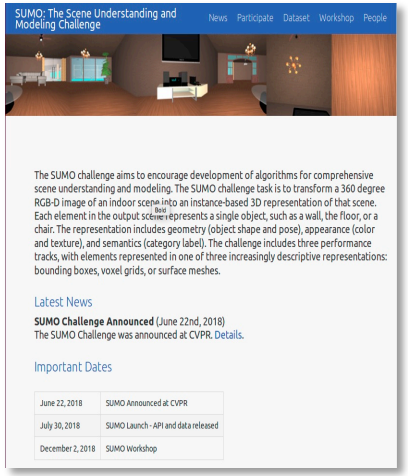
Prizes

Track	Rank	Prize
Surfaces	1st	$2,500 & Titan Xp GPU
Voxels	2nd	$2,000 & Titan Xp GPU
Bounding Boxes	3rd	$1,500 & Titan Xp GPU

SUMO
THE SCENE UNDERSTANDING & MODELING CHALLENGE

Join the challenge at
https://sumochallenge.org/

Facebook Open Source



SUMO Challenge Thanks

Facebook Team

Jay Huang
Facebook

Nandita Nayak
Facebook

John Princen
Facebook

Bahram Dahi
Facebook

Lyne Tchapmi
Stanford

Frank Dellaert
Georgia Tech

Daniel Huber
Facebook

Advisory Board

Ilke Demir
Facebook

Tom Funkhouser
Princeton

Leo Guibas
Stanford

Jitendra Malik
Berkeley

Silvio Savarase
Stanford

SUMO Challenge Thanks

EvalAI

Rishabh Jain
Georgia Tech

Deshraj Yadav
Georgia Tech