JUNE 18-22, 2023
CVPR
VANCOUVER, CANADA

TUE-AM-074

On the Importance of Accurate Geometry Data for Dense 3D Vision Tasks

HyunJun Jung[*1], Patrick RuhKamp[*1,2], Guangyao Zhai[1], Nikolas Brasch[1], Yitong Li[1], Yannick Verdie[3], Jifei Song[3], Yiren Zhou[3], Anil Armagan[3], Slobodan Ilic[4], Ales Leonardis[3], Nassir Navab[1], Benjamin Busam[1,2]

hyunjun.jung@tum.de , p.ruhkamp@tum.de , b.busam@tum.de

Technical University of Munich [1], 3Dwe [2], Huawei Noah's Ark [3], Siemens [4], Equal Contribution [*]

Technical University of Munich

3Dwe

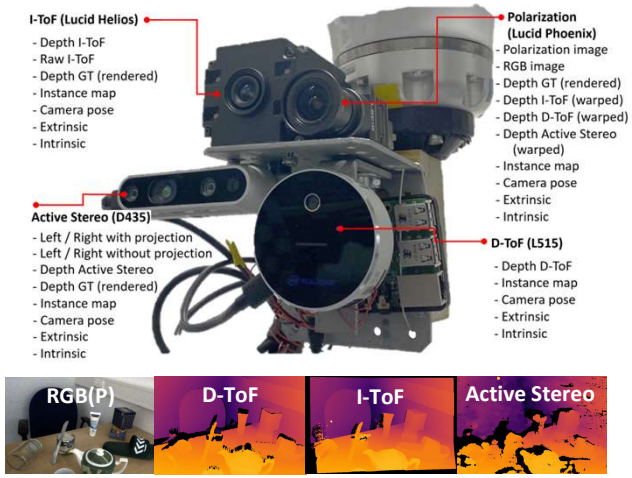
HUAWEI

SIEMENS

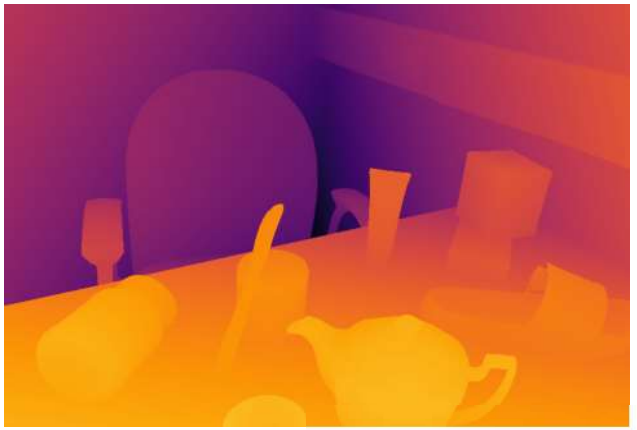
Overview : Intro

HAMMER dataset : Highly Accurate Multi-Modal Dataset for Dense 3D Scene Regression

Full Mesh Scanning with
Accurate Robotic Annotation



Multimodality (RGB + P + 3 x Depth)



High Quality Rendered Depth GT

Overview : Dataset Comparison

ScanNet [1] **Matterport3D** [2] **Replica Dataset** [3] HAMMER Dataset

Other Datasets :

Incomplete mesh, not capable of rendering depth GT

VS

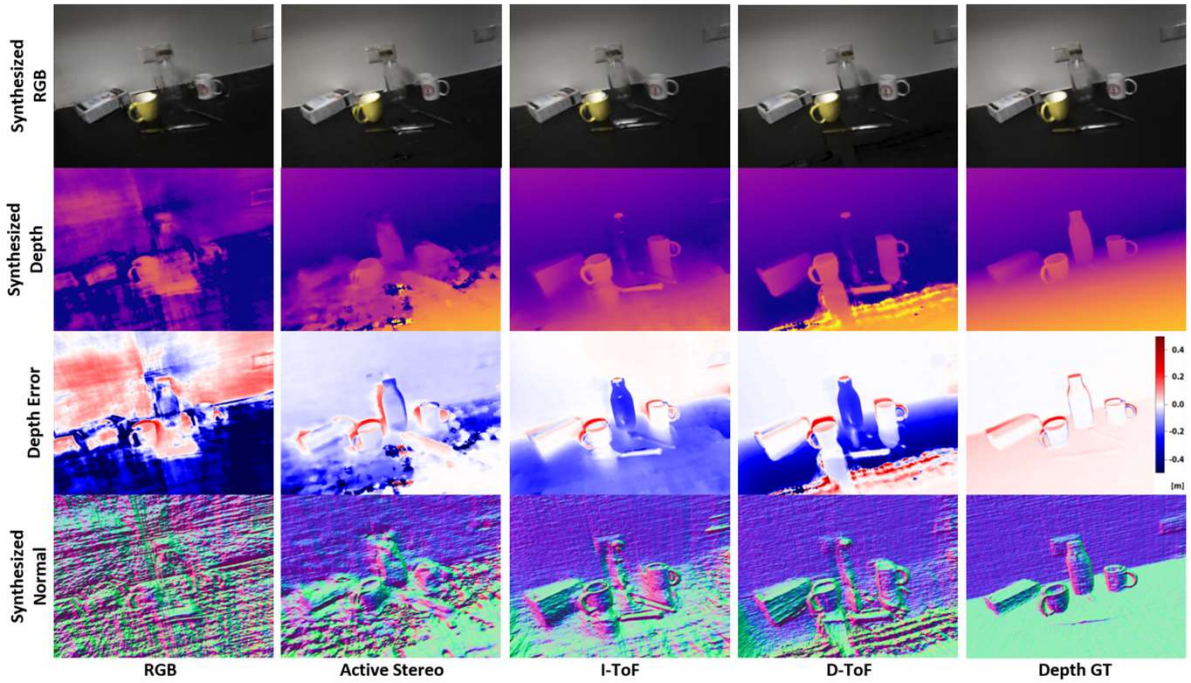
Ours :

Full mesh, capable of rendering depth GT

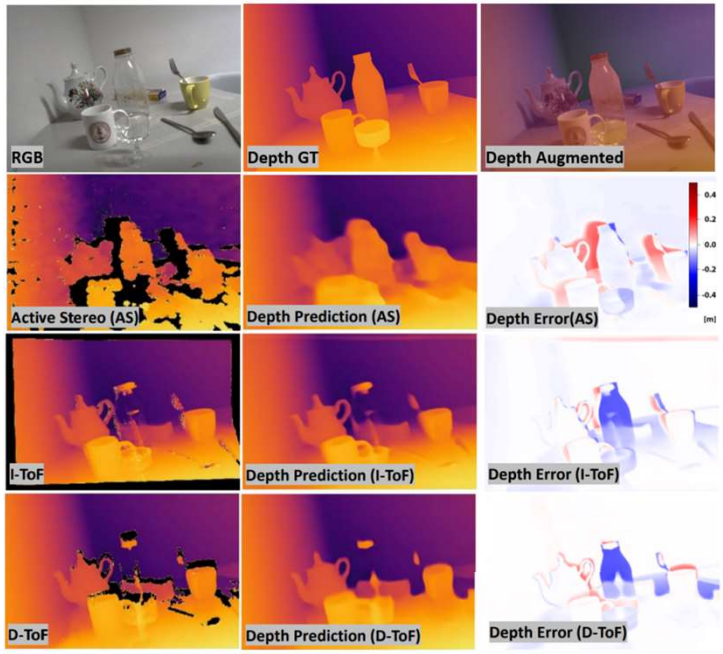
[1] A.Dai, A.Chang, M.Savva, M.Halber, T.Funkhouser, M.Niessner, **ScanNet : Richly-annotated 3D Reconstructions of Indoor Scenes (CVPR 2017)**
[2] A.Chang, A.Dai, T.Funkhouser, H.Halber, M.Niessner, M.Savva, S.Song, A.Zeng, Y.Zhang, **Matterport3D : Learning from RGB-D Data in Indoor Environments (3DV 2017)**
[3] J.Straub, T.Whelan, L.Ma, Y.Chen, E.Wijmans, S.Green, J.J.Engel, R.Mur-Artal, C.Ren, S.Verma, A.Clarkson, M.Yan, B.Budge, Y.Yan, X.Pan, J.Yon, Y.Zou, K.Leon, N.Carter, J.Briales, T.Gillingham, E.Mueggler, L.Pesqueria, M.Savva, D.Batra, H.M.Strasdat, R.D.Nardi, M.Goesele, S.Lovegrove, R.Newcombe **The Replica Dataset : Digital Replica of Indoor Spaces (arXiv 2019)**

Overview : Mainstream 3D Vision Experiments

Vanilla [1]**, Depth Supervised NeRF** [2]



Monocular Depth Estimation [3]



Only dataset that enables direct comparison between different depth modality with perfect ground truth!

[1] B.Mildenhall, P.P.Srinvasan, M.Tancik, J.T.Barron, R.Ramamoorthi, R.Ng – **NERF: Representing Scenes as Neural Radiance Fields for View Synthesis (ECCV 2020)**
[2] B.Roessle, J.T.Barron, B.Mildenhall, P.P.Srinvasan, M.Niessner - **Dense Depth Prior for Neural Radiance Fields from Sparse Input Views (CVPR 2022)**
[3] C.Godard, O.Aodha, M.Firman, G.Brostow, **Digging into self-supervised monocular depth prediction (ICCV 2019)**

1. Dataset Acquisition

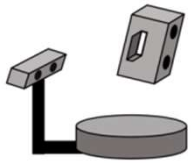
2. Experiments

1. Dataset Acquisition

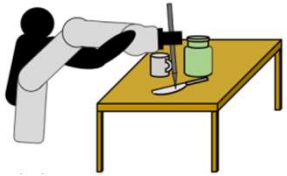
2. Experiments

Dataset Acquisition Overview

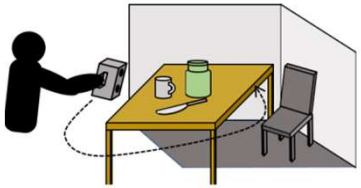
Mesh Scanning and Annotation



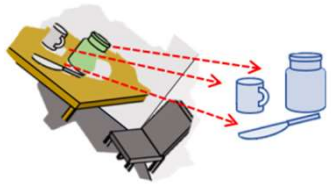
(a) 3D Scanning



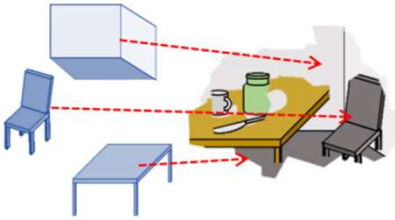
(b) Object Annotation



(c) Partial Scanning

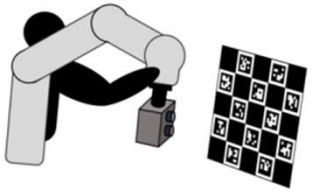


(d) Partial Scanning Fitting



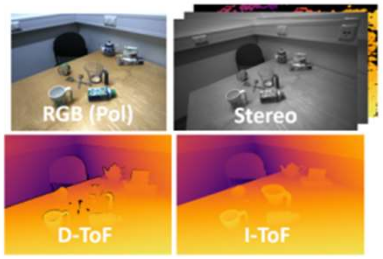
(e) Room / Furniture Fitting

Trajectory Recording and Depth Rendering



(a) Hand-Eye-Calibration

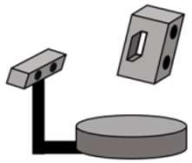
(b) Trajectory Recording



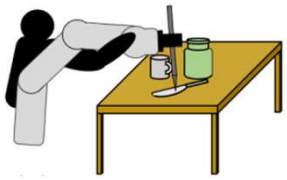
(c) Sequence Recording

(e) Depth Rendering from annotated Mesh

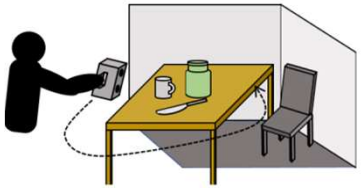
Mesh Scanning and Annotation



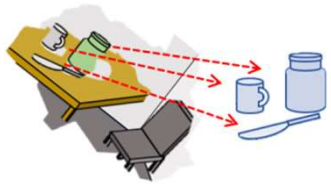
(a) 3D Scanning



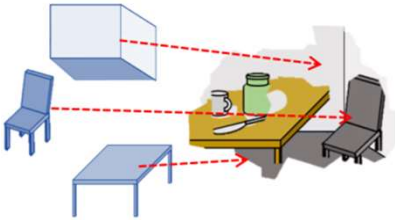
(b) Object Annotation



(c) Partial Scanning



(d) Partial Scanning Fitting

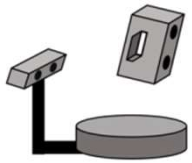


(e) Room / Furniture Fitting

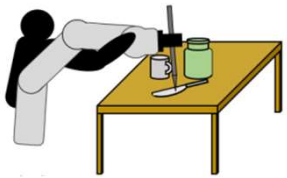
Shinning 3D einscan for small objects

Artec Eva for large objects & room

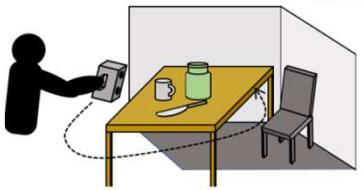
Mesh Scanning and Annotation



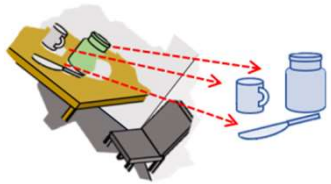
(a) 3D Scanning



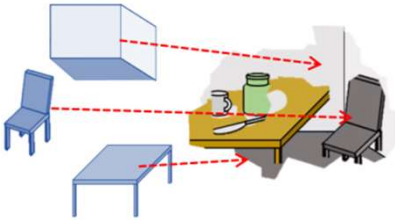
(b) Object Annotation



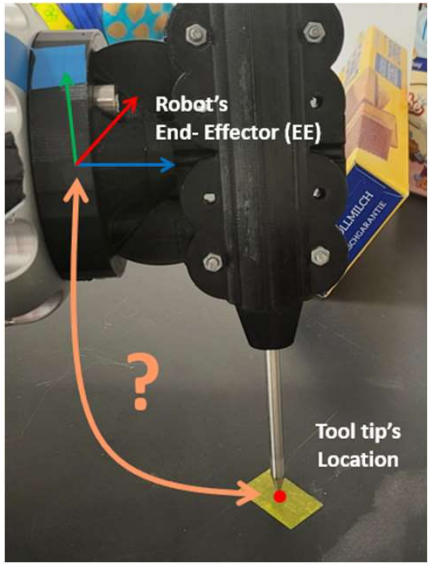
(c) Partial Scanning



(d) Partial Scanning Fitting



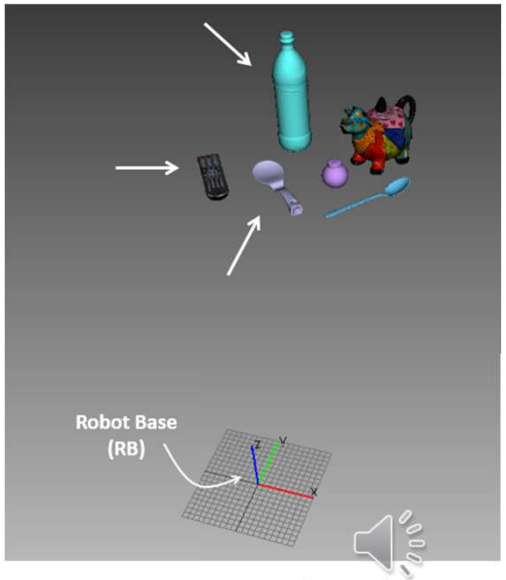
(e) Room / Furniture Fitting



Pivot Calibration

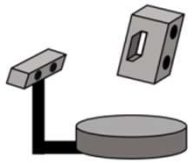
...

Object Surface Measurement

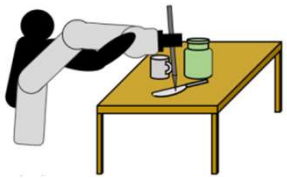


Pose Annotation via correspondence and ICP

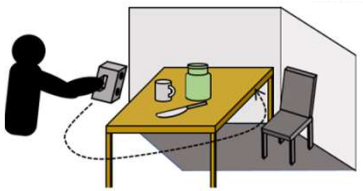
Mesh Scanning and Annotation



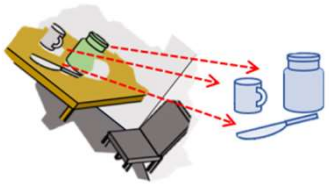
(a) 3D Scanning



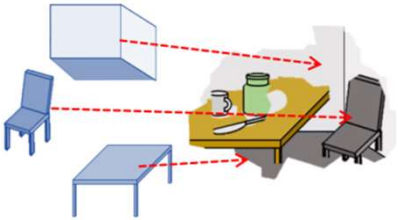
(b) Object Annotation



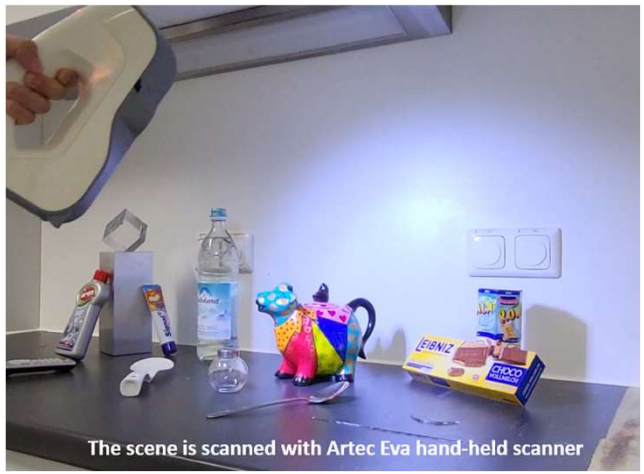
(c) Partial Scanning



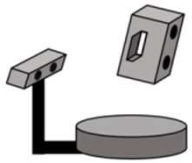
(d) Partial Scanning Fitting



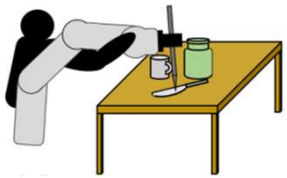
(e) Room / Furniture Fitting



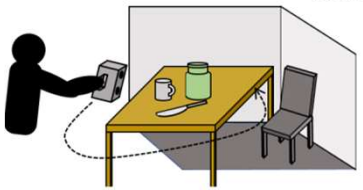
Mesh Scanning and Annotation



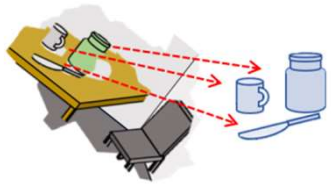
(a) 3D Scanning



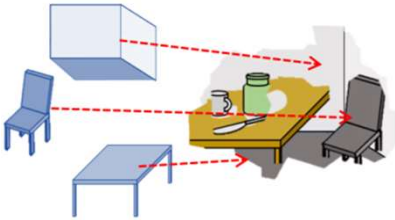
(b) Object Annotation



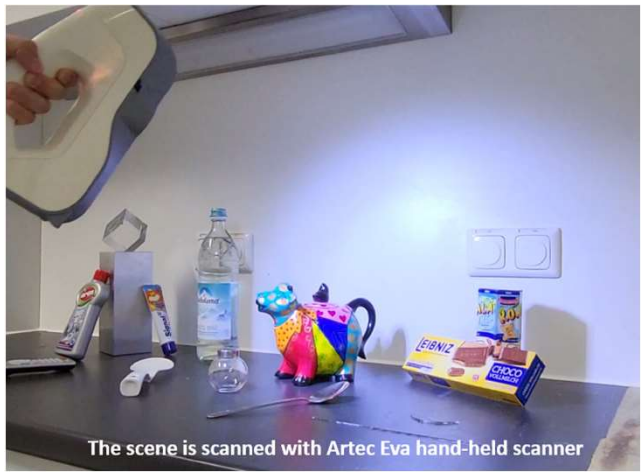
(c) Partial Scanning

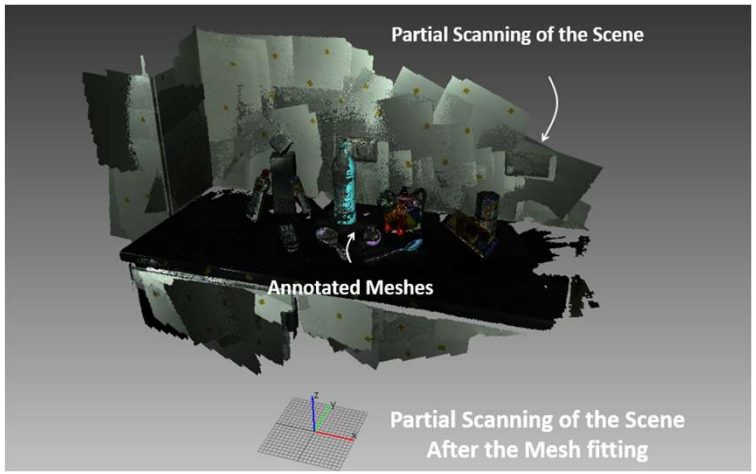


(d) Partial Scanning Fitting

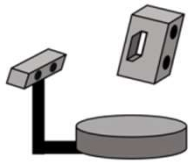


(e) Room / Furniture Fitting

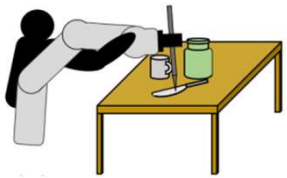




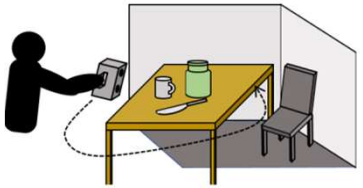
Mesh Scanning and Annotation



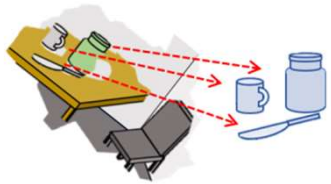
(a) 3D Scanning



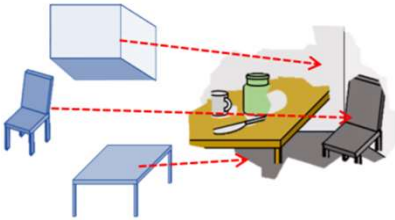
(b) Object Annotation



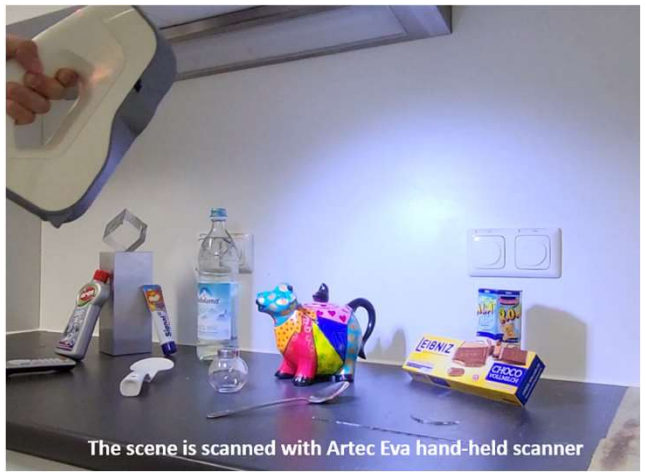
(c) Partial Scanning

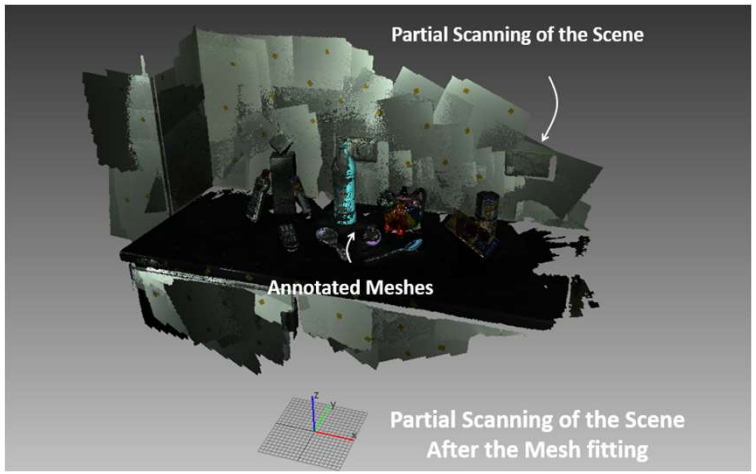


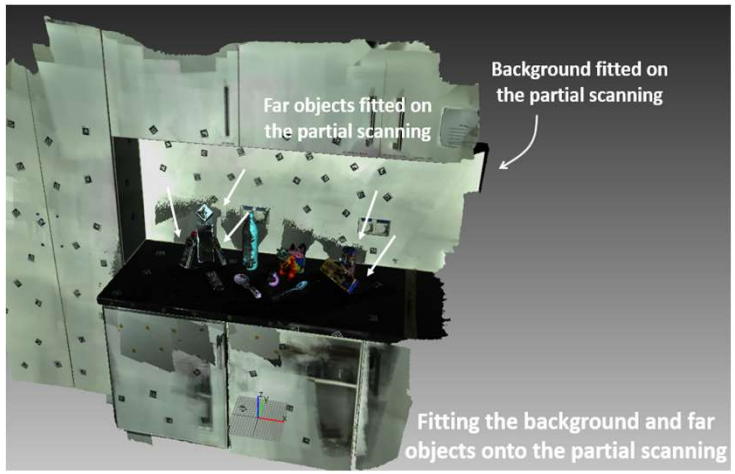
(d) Partial Scanning Fitting



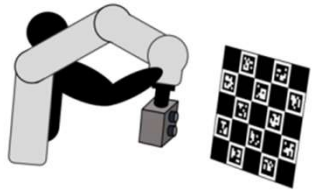
(e) Room / Furniture Fitting



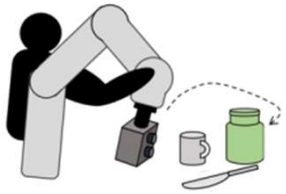




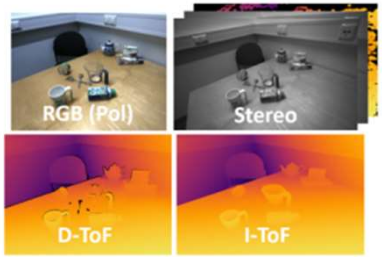
Trajectory Recording and Depth Rendering



(a) Hand-Eye-Calibration

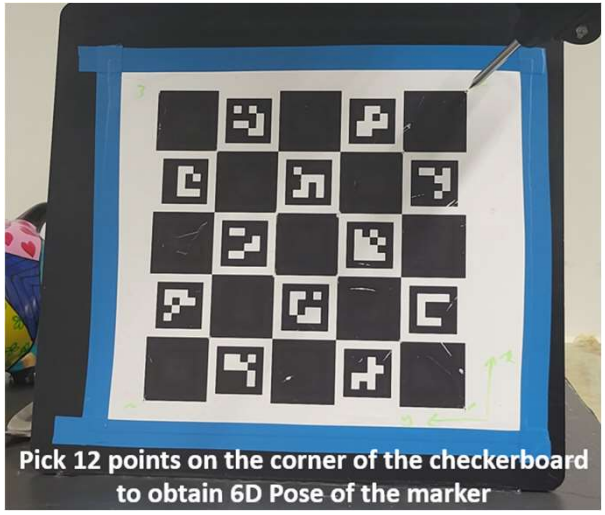


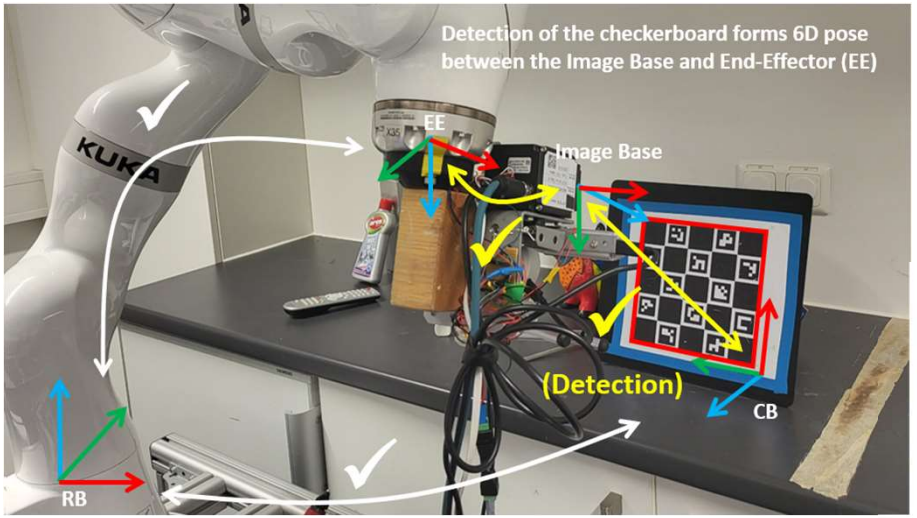
(b) Trajectory Recording



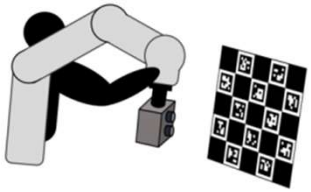
(c) Sequence Recording

(e) Depth Rendering from annotated Mesh

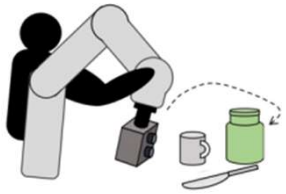




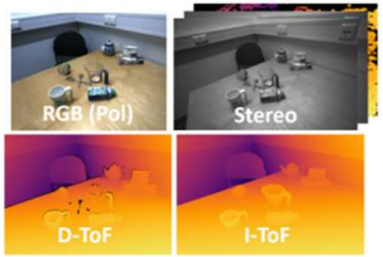
Trajectory Recording and Depth Rendering



(a) Hand-Eye-Calibration



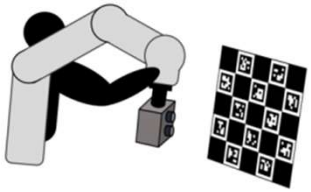
(b) Trajectory Recording



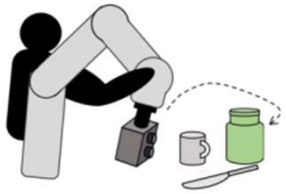
(c) Sequence Recording

(e) Depth Rendering from annotated Mesh

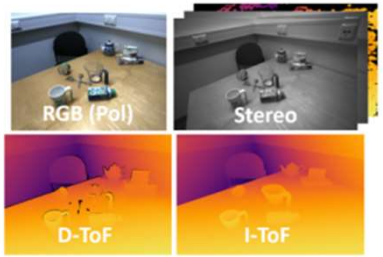
Trajectory Recording and Depth Rendering



(a) Hand-Eye-Calibration



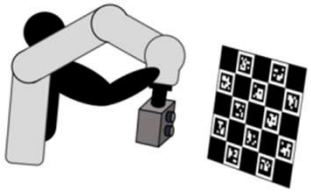
(b) Trajectory Recording



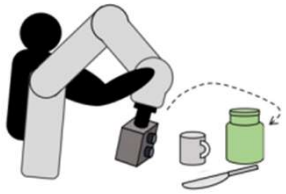
(c) Sequence Recording

(e) Depth Rendering from annotated Mesh

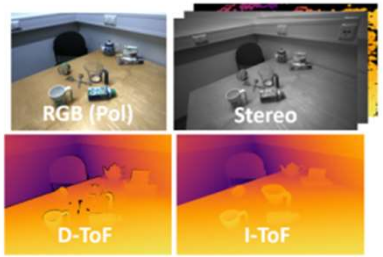
Trajectory Recording and Depth Rendering



(a) Hand-Eye-Calibration



(b) Trajectory Recording



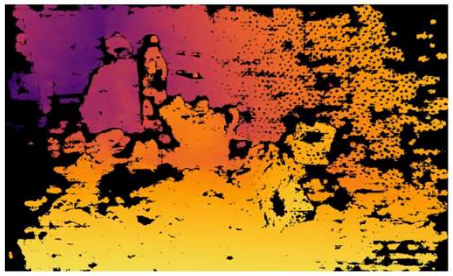
(c) Sequence Recording

(e) Depth Rendering from annotated Mesh

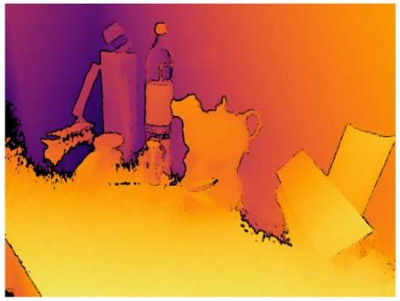
RGB (Polarization)

Instance Mask Augmentation

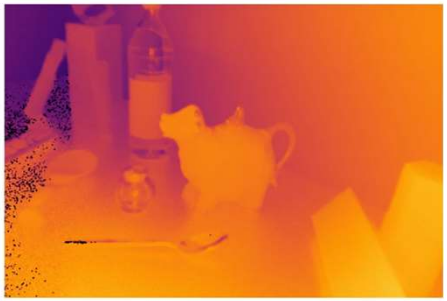
Rendered Ground Truth Depth



Active Stereo Depth



D-ToF (LiDAR) Depth

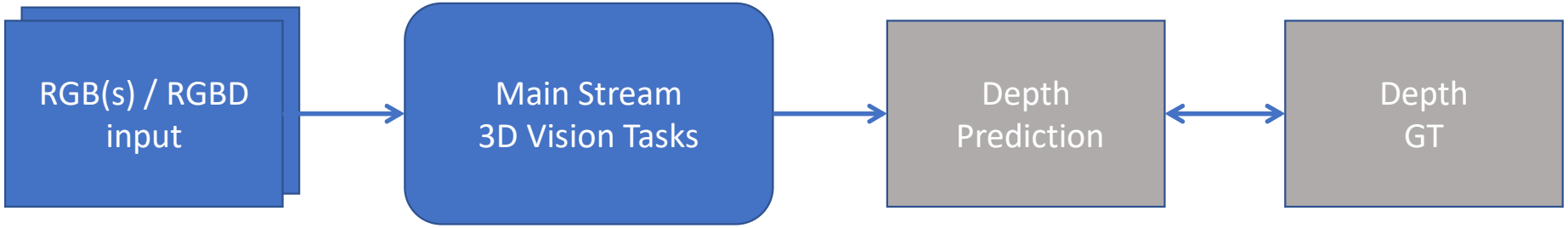


I-ToF Depth

1. Dataset Acquisition

2. Experiments

Experiments Overview

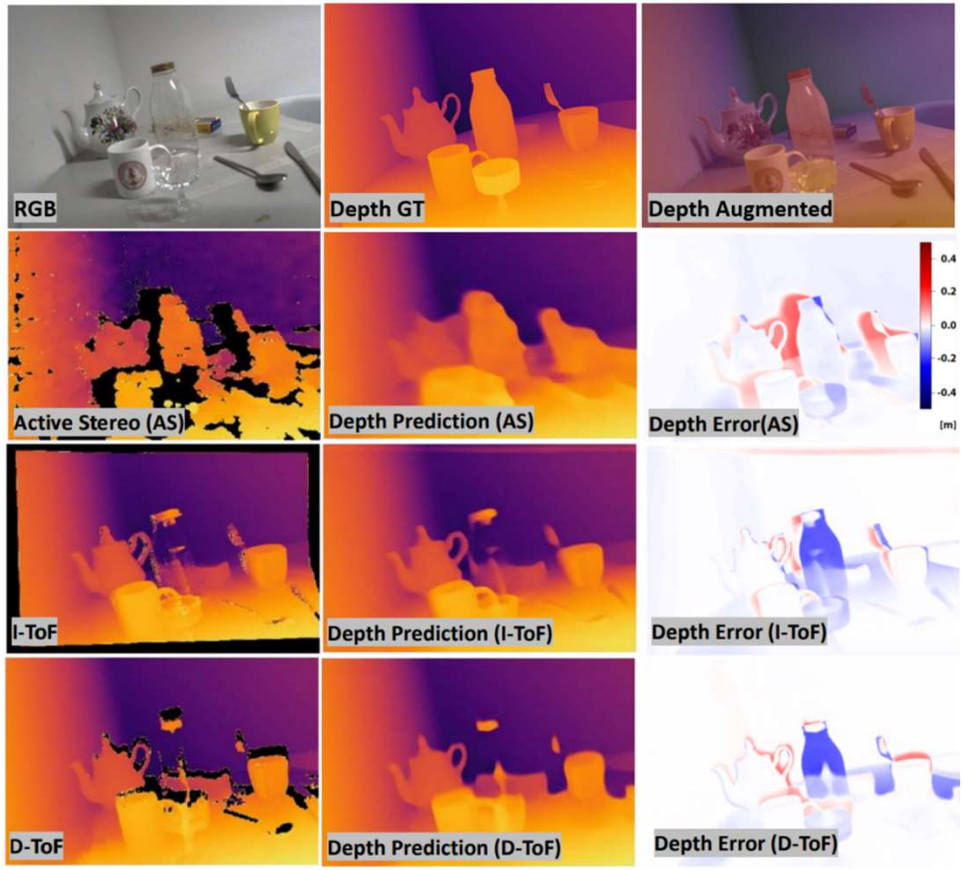


- Monocular Depth Estimation
- RGB(D) NeRF
- Depth Fusion
- Depth Completion

…

- Self Supervision
- Active Stereo Depth
- I-ToF Depth
- D-ToF Depth
- **Rendered GT Depth**

Monocular Depth Estimation



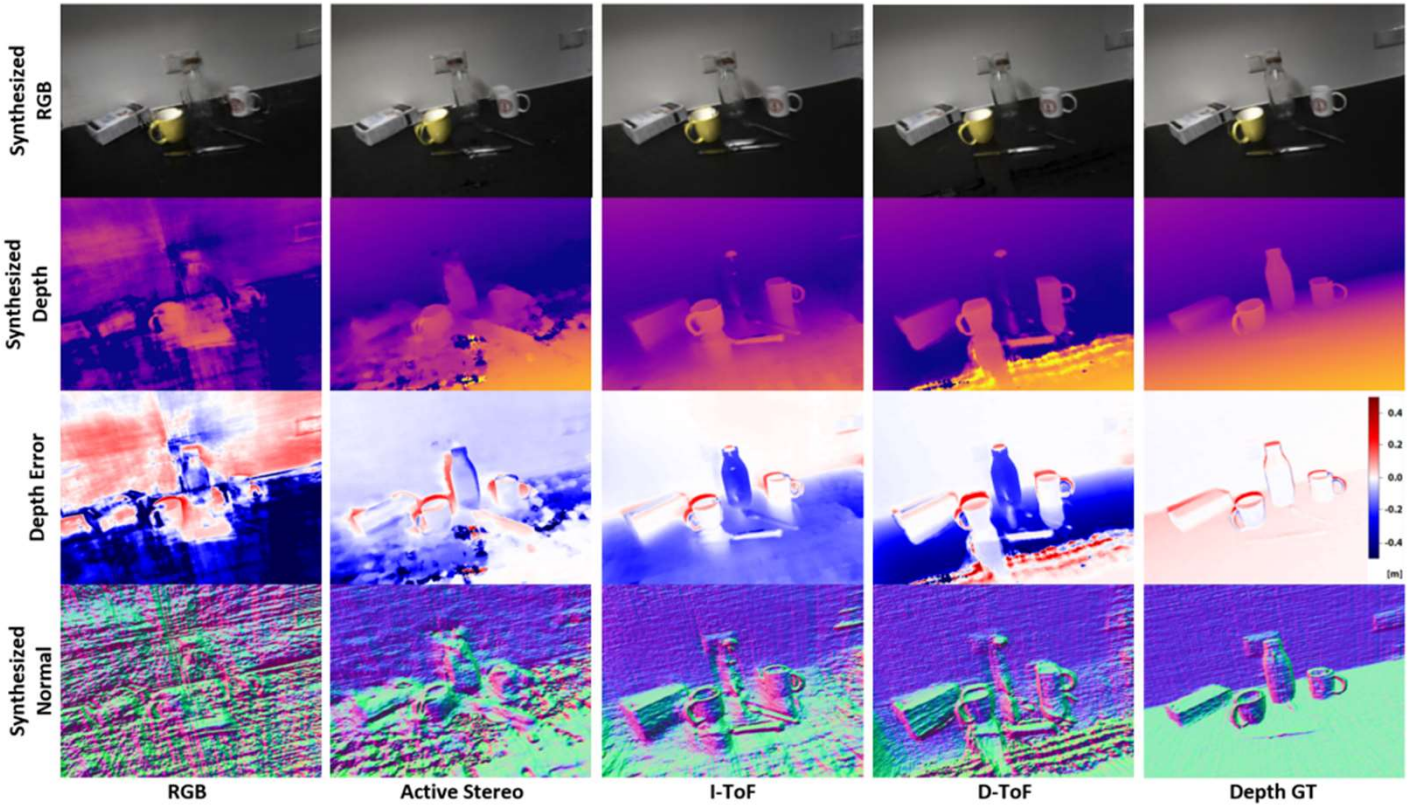
	Training Signal	Full	BG	Obj	Text.	Refl.	Transp.
Sup.	I-ToF	113.29	111.13	119.72	54.45	87.84	207.89
	D-ToF	77.97	**69.87**	112.83	**37.88**	71.59	207.85
	Active Stereo	**72.20**	71.94	**61.13**	50.90	**52.43**	**87.24**
Sel/Sem	Pose	**154.87**	**158.67**	**65.42**	**57.22**	**37.78**	61.86
	M	180.34	183.65	85.51	84.26	48.80	**49.62**
	M+S	159.80	161.65	82.16	71.24	63.92	66.48

RMSE(mm)

Supervised : **MonoDepth2** [1] **with supervision**
Self-Supervised : **MonoDepth2** [1] **with self-supervision**

[1] C.Godard, O.Aodha, M.Firman, G.Brostow, **Digging into self-supervised monocular depth prediction (ICCV 2019)**

NeRF with Depth Prior

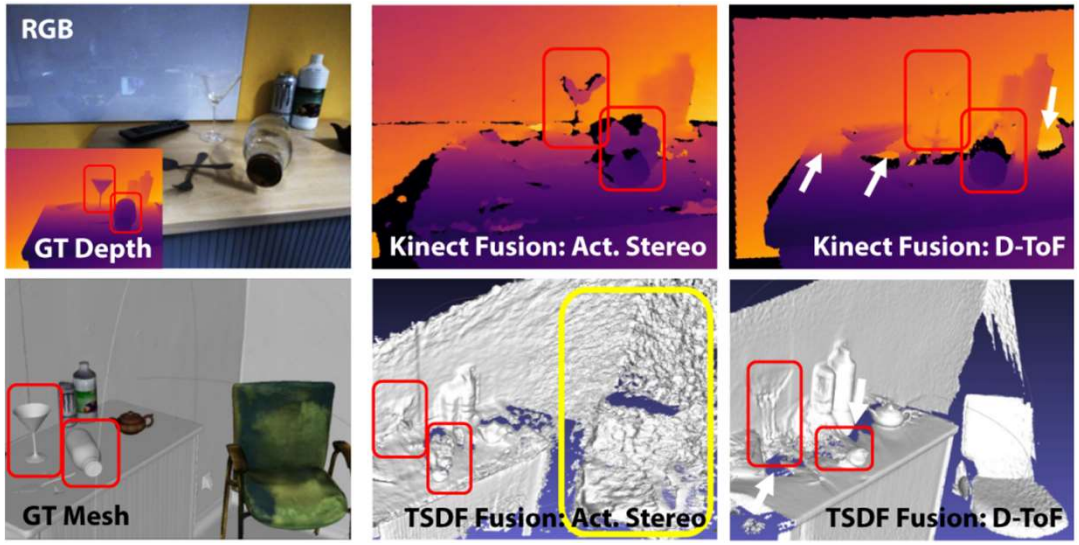


Modality	RGB PSNR ↑	RGB SSIM ↑	Depth Abs.Rel.↓	Depth Sq.Rel.↓	Depth RMSE ↓	Depth $\sigma < 1.25$ ↑
RGB Only	**32.406**	0.889	0.328	111.229	226.187	0.631
+ AS	17.570	0.656	0.113	16.050	94.520	0.853
+ I-ToF	18.042	0.653	0.296	91.426	217.334	0.520
+ D-ToF	31.812	0.888	0.112	24.988	119.455	0.882
+ Syn.	32.082	**0.894**	**0.001**	**0.049**	**3.520**	**1.000**

RGB Only : **NeRF** [1]
RGB + Depth : **Dense Depth Prior NeRF** [2]

[1] B.Mildenhall, P.P.Srinvasan, M.Tancik, J.T.Barron, R.Ramamoorthi, R.Ng – **NERF: Representing Scenes as Neural Radiance Fields for View Synthesis (ECCV 2020)**
[2] B.Roessle, J.T.Barron, B.Mildenhall, P.P.Srinvasan, M.Niessner - **Dense Depth Prior for Neural Radiance Fields from Sparse Input Views (CVPR 2022)**

Sensor Fusion



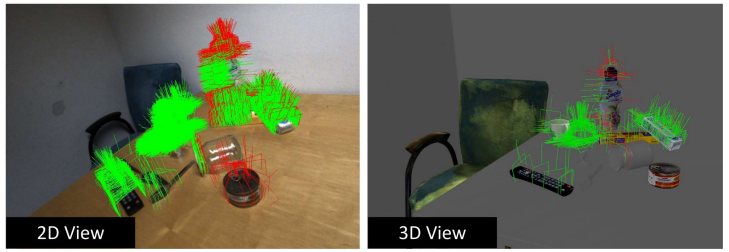
Multi View Depth Fusion **(TSDF** [1] **, Kinect Fusion** [2]**)**

RGBD Fusion **(Wild-ToFu** [3]**)**

[1] Q.Zhou, J.Park, V.Koltun, **Open3D: A modern library for 3D data processing (arxiv 2018)**
[2] R.Newcombe, S.Izadi, O.Hilliges, D.Molyneaux, D.Kim, A.Davison, P.Kohi, J.Shotton, S.Hodges, A.Fitzgibbon, **Kinecfusion: Real-time dense surface mapping and tracking (ISMAR 2011)**
[3] HJ.Jung, N.Brasch, A.Leonardis, N.Navab, B.Busam, **Wild ToFu: Improving range and quality of indirect time-of-flight depth with rgb fusion in challenging environments (3DV 2021)**

Other Possible Applications

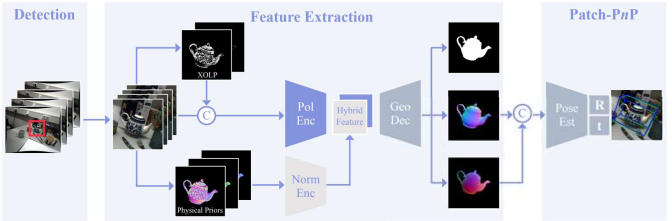
Fully Annotated Meshes
+ Camera Trajectory + Multimodal Camera



Robotic Grasping Annotations [1]

Robotic Grasping Application [1]

Retrieve the Relative Pose
on the Challenging Objects



Pose Estimation Application
on the Challenging Materials [2]

[1] G.Zhai, D.Huang, SC.Wu, HJ.Jung, Y.Di, F.Manhardt, F.Tombari, N.Navab, B.Busam, **MonoGraspNet: 6-DoF Grasping with a Single RGB Image (ICRA 2023)**
[2] D.Gao, Y.Li, P.Ruhkamp, I.Skobelva, M.Wysocki, HJ.Jung, P.Wang, A.Guridi, B.Busam, **Polarimetric Pose Prediction (ECCV 2022)**

JUNE 18-22, 2023
CVPR
VANCOUVER, CANADA

TUE-AM-074

On the Importance of Accurate Geometry Data for Dense 3D Vision Tasks

HyunJun Jung[*1], Patrick RuhKamp[*1,2], Guangyao Zhai[1], Nikolas Brasch[1], Yitong Li[1], Yannick Verdie[3], Jifei Song[3], Yiren Zhou[3], Anil Armagan[3], Slobodan Ilic[4], Ales Leonardis[3], Nassir Navab[1], Benjamin Busam[1,2]

hyunjun.jung@tum.de , p.ruhkamp@tum.de , b.busam@tum.de

Technical University of Munich [1], 3Dwe [2], Huawei Noah's Ark [3], Siemens [4], Equal Contribution [*]

Technical University of Munich

HUAWEI

SIEMENS