DLODepth: Real-time Depth Recovery for 3D Reflective Deformable Linear Object

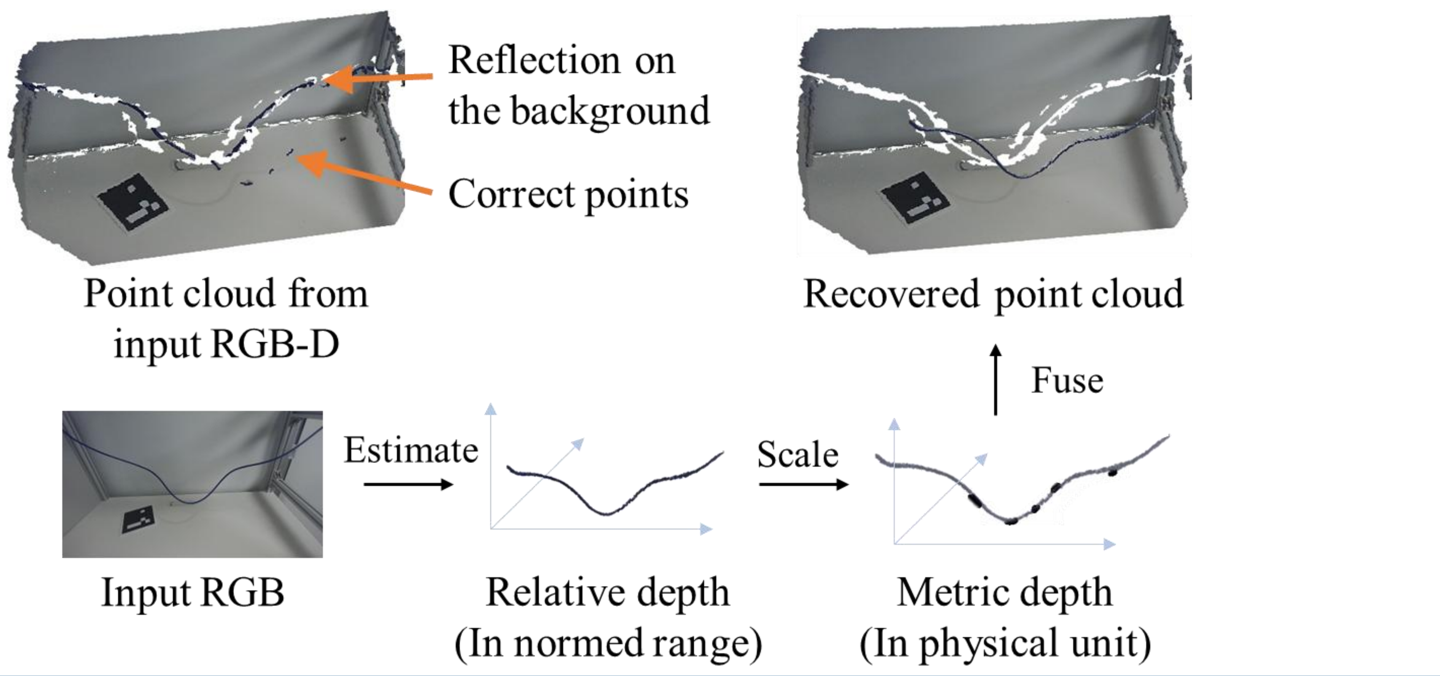
5th workshop on RObotic MAnipulation of Deformable Objects: holistic approaches and challenges forward

Li Huang, Tong Yang, Xiang Tian, Rongxin Jiang, Yaowu Chen
Zhejiang University

Motivation

The depth data of Deformable Linear Object (DLO) is severely affected by geometrical and optical errors due to its thin diameter and reflective surface.

The proposed method recovers DLO point cloud from noisy input monocular RGBD data.



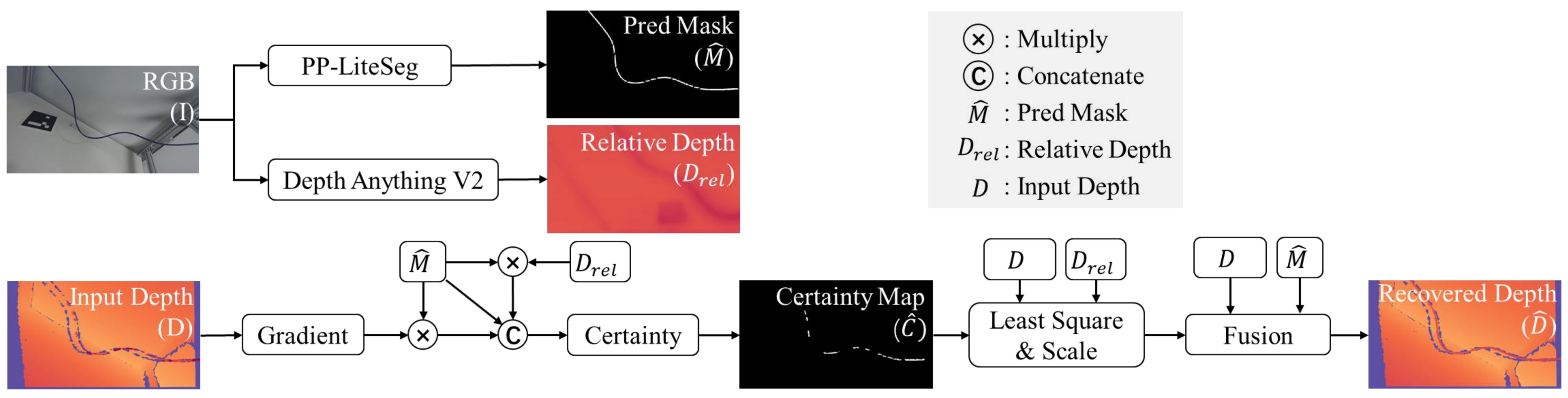
Contribution

1. The first end-to-end framework that directly recovers a spatially accurate 3D shape of a DLO from flawed RGB-D input.
2. SOTA results on real world challenging DLO with real-time inference (0.031s/frame): mean distance error of 4.3cm (1.4cm median), mean recovery rate of 69.4% (93.8% median).
3. A modified distance loss term that compensates for the discrepancy between the pin-hole camera model and Euclidean space.
4. Open-sourced codes, the dataset, and the data collection tool with a GUI front-end.

Algorithm

The proposed algorithm is four-folded:

1. RGB Segmentation (PP-LiteSeg)
2. Relative Depth Estimation (Depth Anything V2)
3. Relative-to-Metric Scaling Transformation
   - Select Anchor
   - Calculate scaler
4. Recovery Fusion



The Three-stage training strategy:

1. Relative Depth (SiLog, Proj, Cont)
2. Anchor Points (Focal)
3. Segmentation (CE)

Loss $\tilde{d} = d \cdot \| \mathbf{n}(u,v) \|_2$

- Projection loss: the mean 3D distance error on the foreground pixels of predictive and GT mask.
- Continuity loss: the local standard deviation predictive depth on foreground pixels of pred mask.

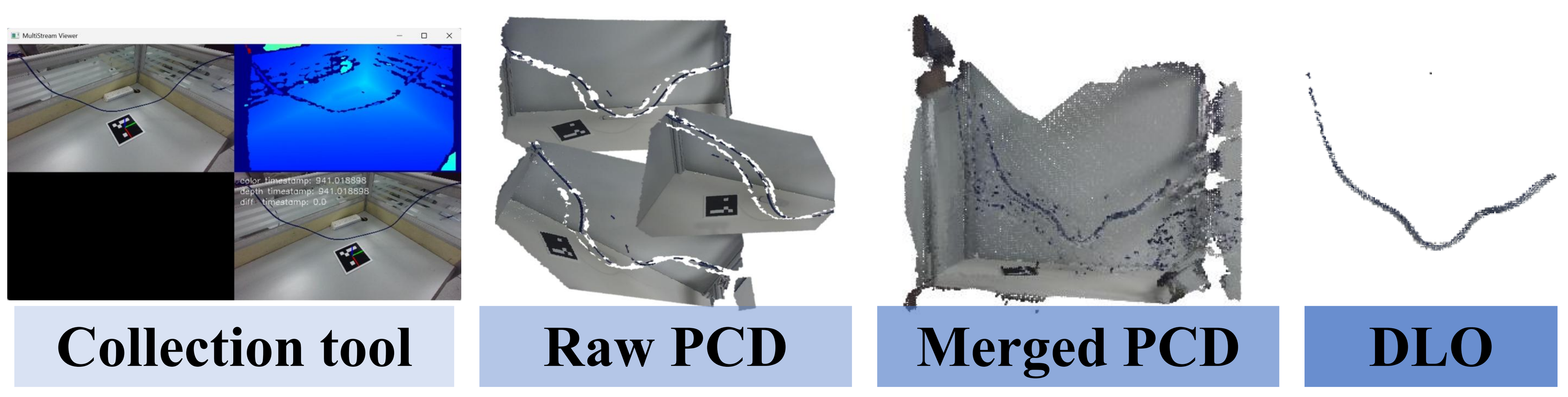
Metric:

- Mean Accuracy Distance ($pred \rightarrow GT$ distance). CR@5cm (Rate of GT with pred within 5cm).

Open-sourced Dataset

DLO dataset is captured in real world.

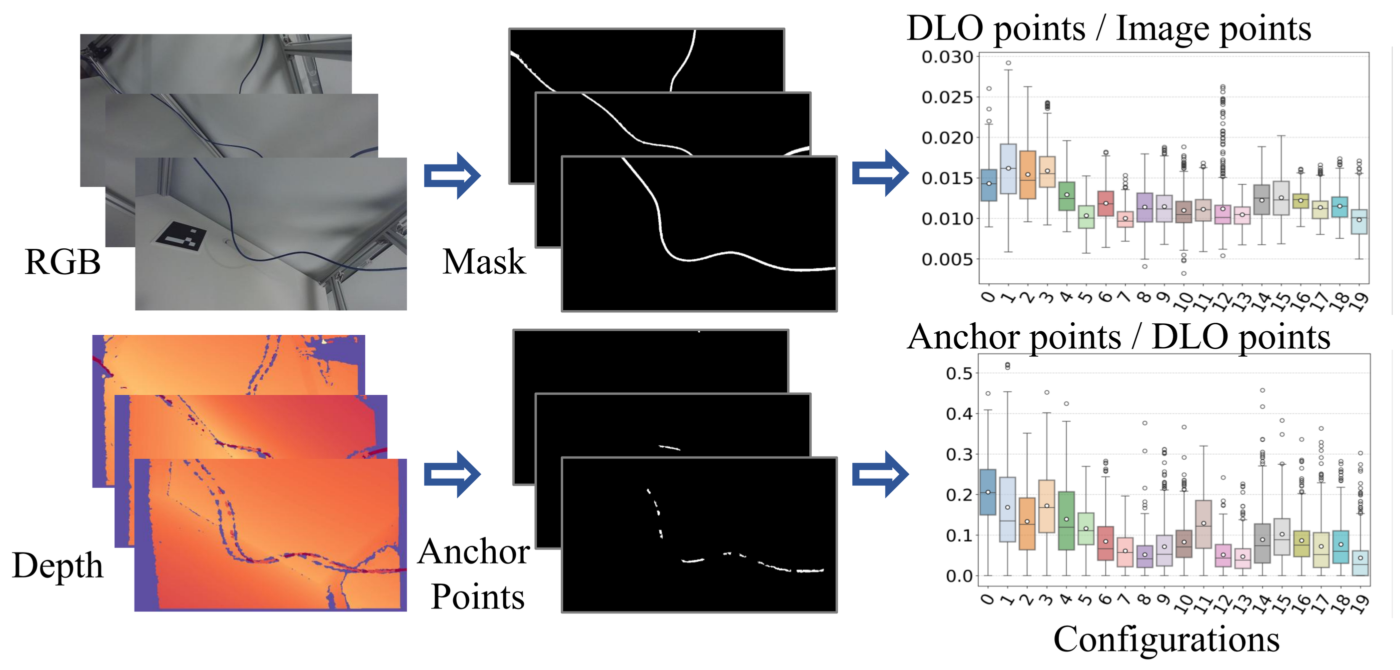
- Handheld monocular RGB-D camera.
- 20 configurations of both blue and yellow Ethernet cables (diameter: 0.6cm).
- 6,840 and 4,636 RGB-D pairs.



- The generated pseudo-GT depth image D^* is

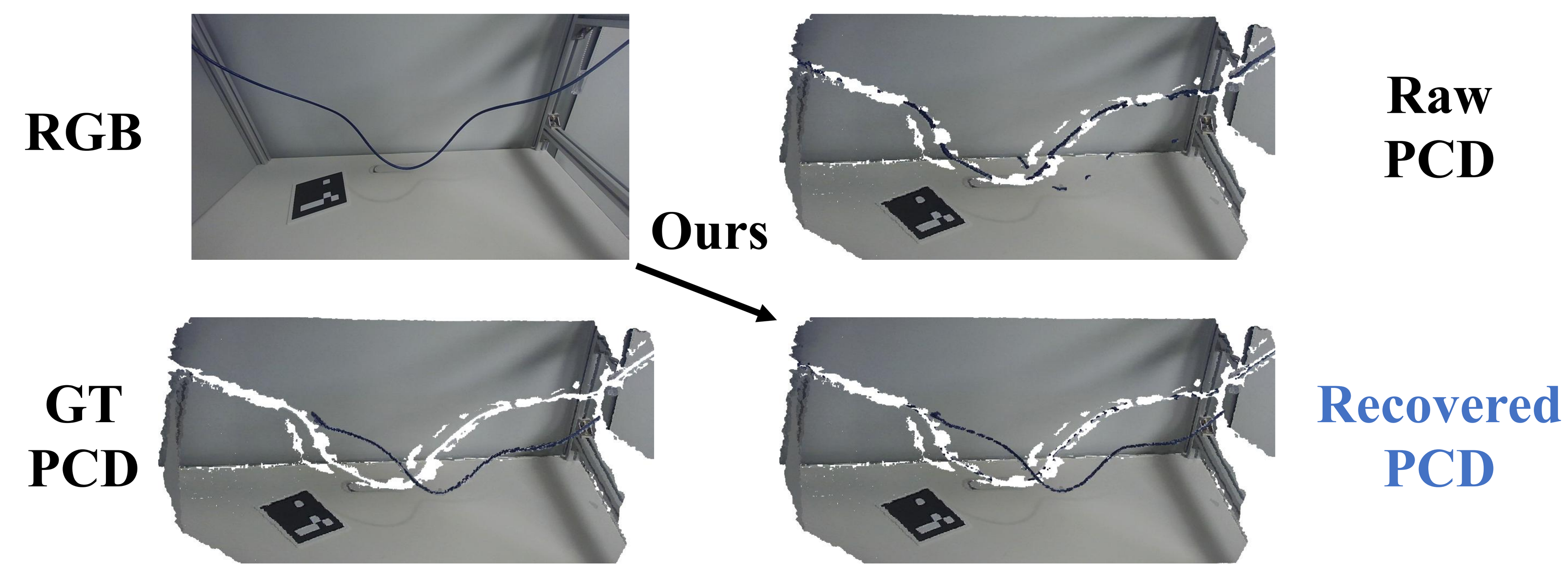
$$D^*(u,v) = \begin{cases} D_{\text{proj}}(u,v), & (u,v) \in \Omega^* \\ D(u,v), & \text{otherwise} \end{cases}$$

- Only 1.5% image pixels belong to DLO. Only 0.3% image pixels are DLO pixels with reliable depth.



Experiment

Visualization of our proposed method.



- The proposed method achieves SOTA and is the only one with MAD less than 5cm.

Methods	Train	Depth	Fusion	Success Num	MAD(m) ↓ mean	MAD(m) ↓ median	CR@5cm(%) ↑ mean	CR@5cm(%) ↑ median	AbsRel ↓	RMSE ↓	δ_1 ↑	Time (s) ↓
Raw depth	-	Met.[1]	-	0(0.0%)	0.310	0.305	48.5	47.6	0.009	0.045	0.988	-
AdaBins [18]	Train	Met.	-	230 (17.57%)	0.081	0.075	66.7	71.9	0.107	0.104	0.904	0.027
AdaBins [18]	Train	Met.	SAM 2 [19]	230 (17.57%)	0.081	0.075	66.7	71.9	0.003	0.016	0.995	0.027+0.143[2]
Metric3Dv2 [20]	Zero[3]	Met.	-	17 (1.30%)	0.351	0.341	2.9	0.0	0.241	0.178	0.644	0.050
Depth Pro [21]	Zero	Met.	-	155 (11.84%)	0.140	0.123	25.1	18.2	0.140	0.126	0.805	0.404
UniDepthV2 [22]	Zero	Met.	-	0(0.00%)	0.542	0.497	0.3	0.0	0.498	0.364	0.270	0.042
Depth Anything V2 [13]	Zero	Rel(lstsq)	-	214 (16.35%)	0.130	0.109	26.6	17.2	0.200	0.181	0.574	0.015
[23](with depthFM [24])	Zero	Met.	completion	1 (0.08%)	0.223	0.220	30.0	23.8	0.044	0.064	0.959	14.64
Ours	Train	Met.	PP-LiteSeg [12]	**957 (73.11%)**	**0.043**	**0.014**	**69.4**	**93.8**	0.005	0.034	0.994	0.031

[1] "Met": metric depth. "Rel": relative depth. "lstsq": least squares fitting.
[2] The 0.143s is the average single frame inference time of Grounded SAM 2 on NVIDIA GeForce RTX 4090 GPU.
[3] "Zero": zero-shot. "Train": fine-tuned on the proposed DLO dataset.

- Robust on fewer training data (train and test for each DLO).

DLO	Train&Val shapes(images)	Test shapes(images)	Success Num	MAD (m) ↓ mean	MAD (m) ↓ median	CR@5cm (%) ↑ mean	CR@5cm (%) ↑ median	AbsRel ↓	RMSE ↓	δ_1 ↑
Blue Cable	16(5531)	4(1309)	957 (73.11%)	**0.043**	0.014	69.4	**93.8**	0.005	0.034	0.994
	10(3472)	10(3368)	**2639 (78.36%)**	0.055	**0.012**	**72.1**	92.8	0.006	0.037	0.992
	5(1599)	15(5241)	3819 (72.87%)	0.065	0.016	65.1	85.9	0.007	0.039	0.992
Yellow Cable	16(3689)	4(947)	**823 (86.91%)**	**0.041**	**0.015**	**76.3**	**87.9**	0.005	0.038	0.994
	10(2371)	10(2265)	1937 (85.52%)	0.049	0.016	73.3	84.9	0.005	0.037	0.994
	5(1138)	15(3498)	2498 (71.41%)	0.077	0.025	60.1	71.2	0.006	0.039	0.994