SDF-2-SDF: Highly Accurate 3D Object Reconstruction

Christian Diller

SDF-2-SDF:
Highly Accurate 3D Object Reconstruction

Miroslava Slavcheva[1,2], Wadim Kehl[1], Nassir Navab[1], Slobodan Ilic[1,2]

[1] Technische Universität München, Munich, Germany
[2] Siemens AG, Munich, Germany
{mira.slavcheva,nassir.navab}@tum.de, {kehl,slobodan.ilic}@in.tum.de

Abstract. This paper addresses the problem of 3D object reconstruction using RGB-D sensors. Our main contribution is a novel implicit-to-implicit surface registration scheme between signed distance fields (SDFs), utilized both for the real-time frame-to-frame camera tracking and for the subsequent global optimization. SDF-2-SDF registration circumvents expensive correspondence search and allows for incorporation of multiple geometric constraints without any dependence on texture, yielding highly accurate 3D models. An extensive quantitative evaluation

Outline

- Motivation
- Previous Approaches
- Proposed Method
- Evaluation
- Conclusion

Motivation

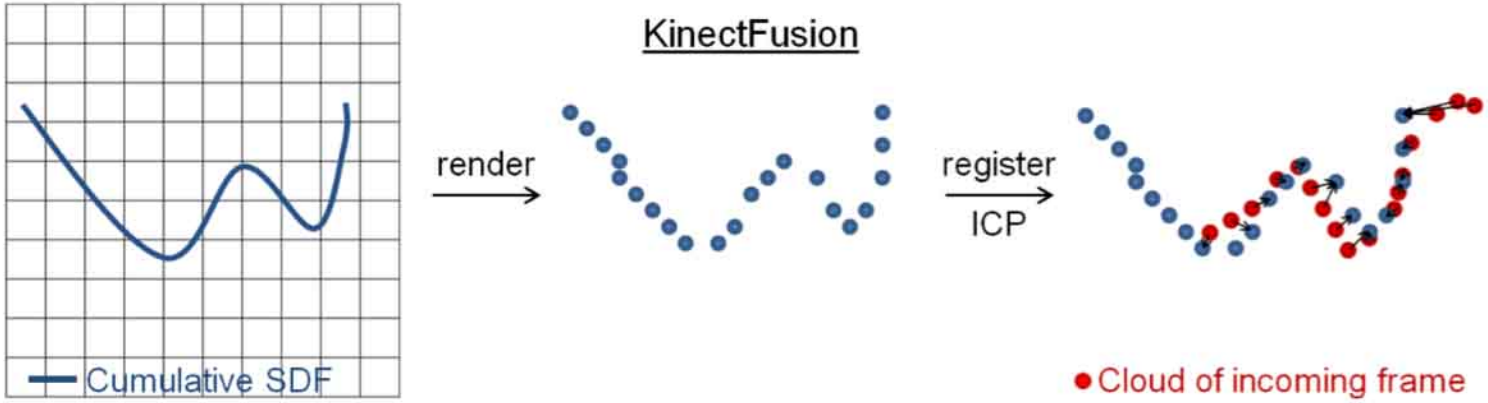
- Most previous solutions are SLAM-like
- Focus on highly accurate 3D object reconstruction
- Avoid expensive explicit correspondence search
- No dependence on texture

Previous approaches

- DVO SLAM
  - Costly graph optimization, might last hours to days
  - Depends on texture in scenes
  - Industrial scanning scenarios might not allow for texturing

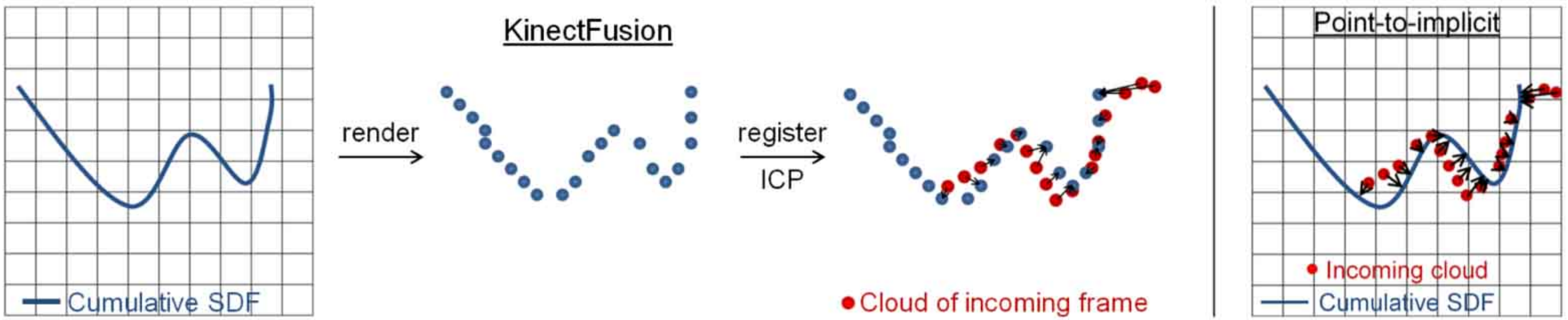
Previous approaches

- KinectFusion
  - ICP cannot handle gross outliers and large motion
  - ICP is costly (recomputing point correspondences)
  - Frame-to-model registration fails when model accumulates error



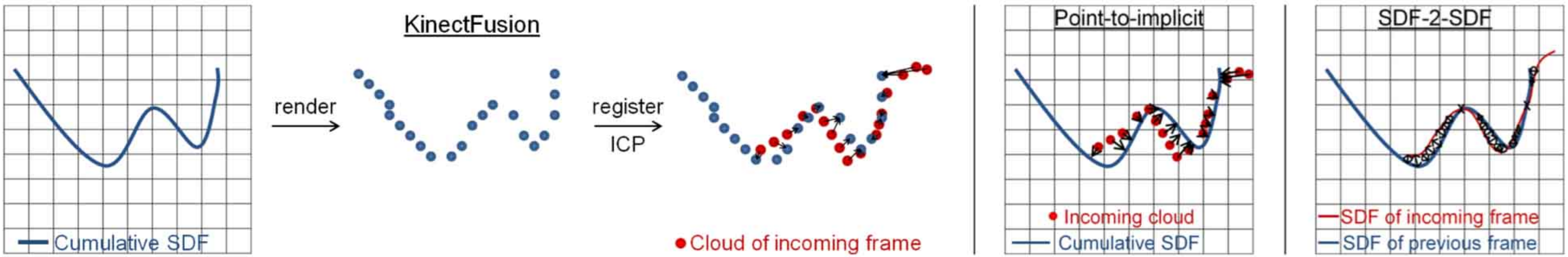
Previous approaches

- Point-to-implicit schemes
  - Directly project incoming depth frames into implicit representation
  - More robust than KinectFusion
  - Unreliable with sparse range data or errors in global model



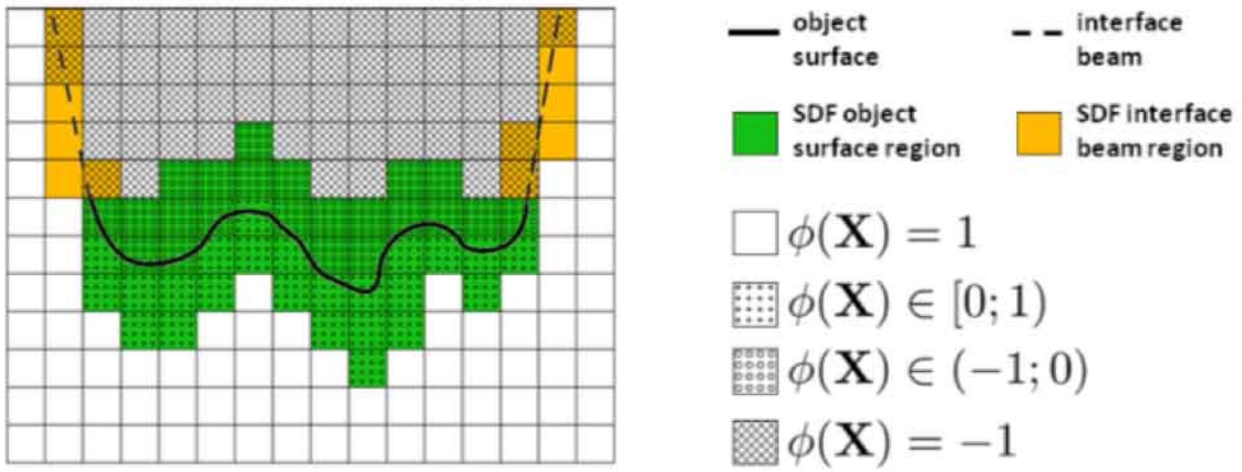
Proposed Method

- Minimize the difference between pairs of SDFs (implicit-to-implicit)
- Online real-time camera tracking and offline global pose refinement
- Photometry independent
- Avoid drift caused by errors in the global model



Signed Distance Function

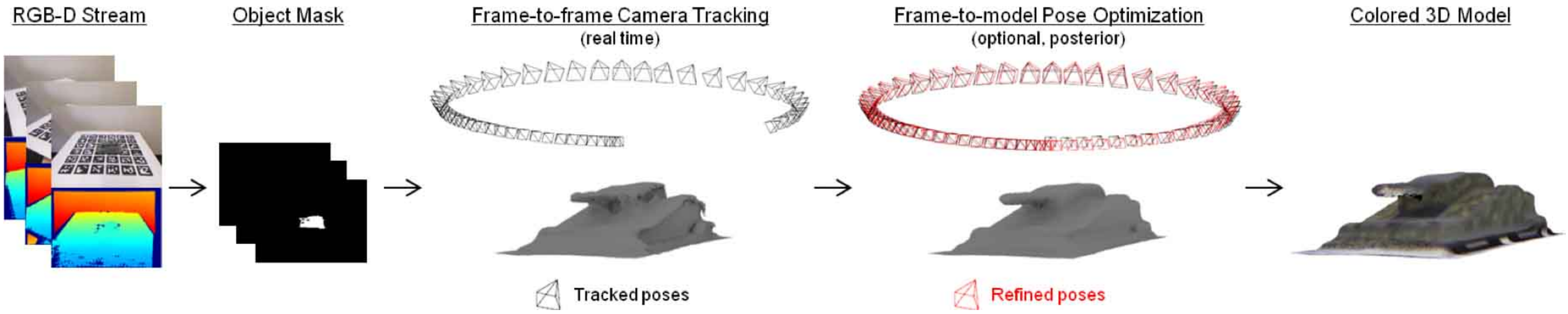
- Same formulation as in KinectFusion



- Smoothing properties that make TSDFs superior to explicit 3D representations

SDF-2-SDF Registration

- Preprocess: Masking, Denoising
- Frame bounding volume estimation
- After tracking, register keyframes into globally weighted average
- Construct colored mesh with Marching Cubes



Objective Function

- Iteratively minimize direct difference between two SDFs
  - SDFs encode the distance to the common surface
  - Same value for truncated voxels, near-surface voxels steer convergence
  - Minimization scheme based on a first-order Taylor approximation

$$E_{SDF}(\xi) = E_{geom}(\xi) + \alpha_{norm} E_{norm}(\xi) \ ,$$

$$E_{geom}(\xi) = \frac{1}{2} \sum_{voxels} \Big(\phi_{ref}\omega_{ref} - \phi_{cur}(\xi)\omega_{cur}(\xi) \Big)^2 \ ,$$

$$E_{norm}(\xi) = \sum_{surface\ voxels} \Big(1 - \overline{\mathbf{n}}_{ref} \cdot (\overline{\mathbf{n}}_{cur}(\xi)) \Big) \ .$$

Global Pose Optimization

- Trajectory might have accumulated drift
- Refine using frame-to-model scheme based on the SDF-2-SDF energy

1. Fuse keyframes into globally weighted SDF
2. Refine all poses
3. Apply pose updates simultaneously
4. Recompute model every few iterations

Evaluation

- Printed 3D objects of varying geometry, size and color
- CAD models, RGB-D data from various sensors and externally measured trajectories

Evaluation

- Qualitative and quantitative comparisons with
  - ICP-based approaches
  - Point-to-implicit methods
  - Visual odometry methods

- Trajectory accuracy
- 3D model accuracy

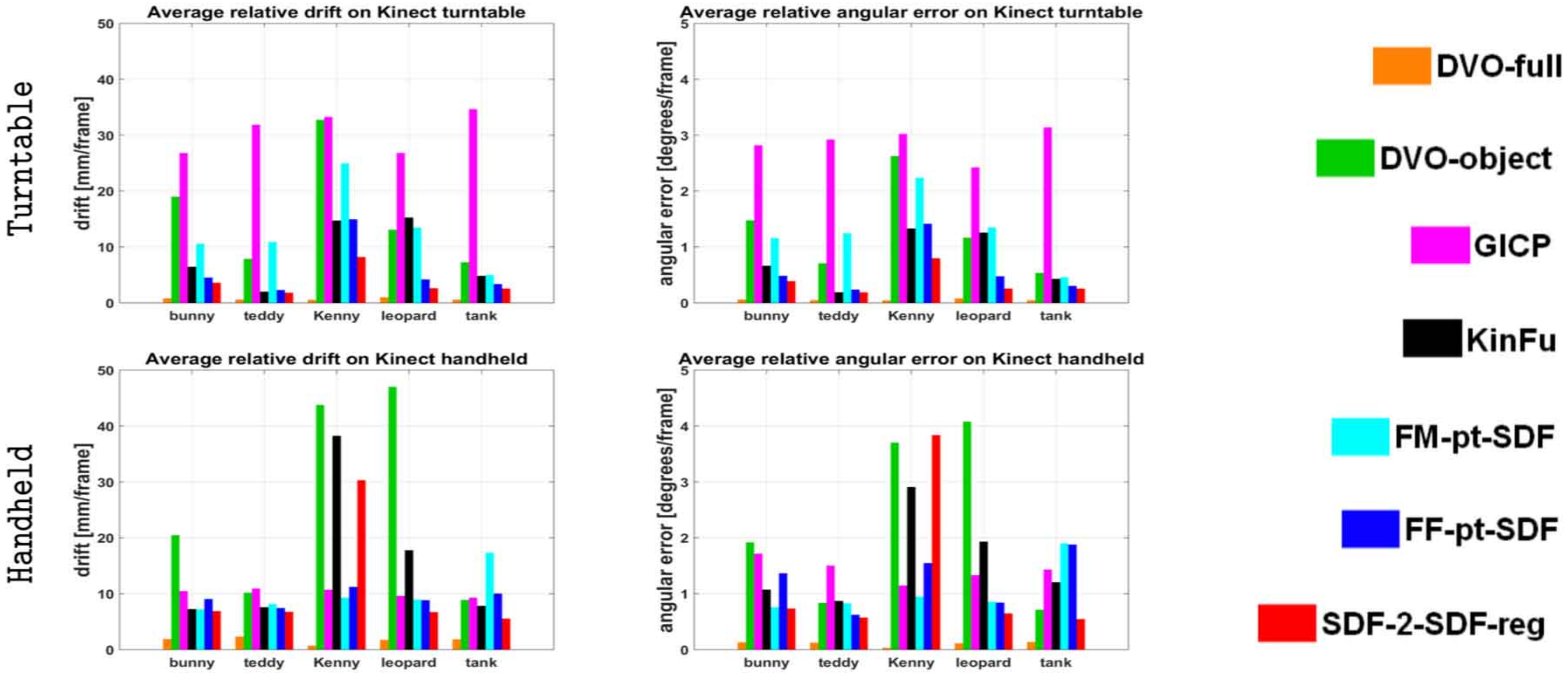
Trajectory accuracy

- Evaluate relative pose error RPE

$$RPE_{i \to i+1} = (\mathbf{P_i}^{-1}\mathbf{P_{i+1}})^{-1}(\mathbf{Q_i}^{-1}\mathbf{Q_{i+1}})$$

- Calculate root-mean-squared, average, minimum and maximum
- Also using RPE for evaluation of absolute poses
- Additionally, measure angular error

Trajectory accuracy



3D model accuracy

- Measuring the absolute distance mean and standard deviation of models

- Compare meshes to
  - KinectFusion without refinement
  - Method of Kehl et al. (DVO, loop closure, graph optimization) with refinement

3D model accuracy

object	method	error [mm]									
		synth. circle		synth. wave		industr. turntab.		Kinect turntable		Kinect handheld	
		mean	std.dev.	mean	std.dev.	mean	std.dev.	mean	std.dev.	mean	std.dev.
bunny	KinFu	0.544	0.677	0.787	0.988	0.664	0.654	3.800	2.840	4.101	3.716
	ours (no refinement)	0.135	0.139	0.133	0.134	0.656	0.438	2.586	1.869	1.770	1.733
	Kehl *et al.* object	1.459	1.220	4.885	3.732	2.149	2.869	5.156	4.115	8.274	6.013
	Kehl *et al.* full	—	—	—	—	0.838	0.860	1.134	1.243	1.124	1.095
	ours (with refin.)	**0.130**	**0.137**	**0.131**	**0.133**	**0.541**	**0.436**	**0.953**	**0.843**	**0.996**	**0.853**
teddy	KinFu	0.370	0.275	0.418	0.285	0.998	0.807	1.271	1.045	2.355	1.447
	ours (no refinement)	0.161	0.179	0.146	0.142	0.930	0.588	1.078	0.890	1.589	1.537
	Kehl *et al.* object	0.358	0.303	0.257	0.193	1.028	0.892	2.306	1.862	2.287	1.826
	Kehl *et al.* full	—	—	—	—	4.828	4.215	1.221	0.858	3.066	2.380
	ours (with refin.)	**0.157**	**0.166**	**0.146**	**0.142**	**0.910**	**0.584**	**0.722**	**0.542**	**0.990**	**0.841**
Kenny	KinFu	0.418	0.311	0.440	0.359	1.650	1.451	1.511	1.387	2.874	2.727
	ours (no refinement)	0.154	0.151	0.147	0.154	0.363	0.391	1.295	1.311	2.415	2.051
	Kehl *et al.* object	0.948	0.736	1.931	1.965	1.816	1.710	3.181	3.238	*failed*	*failed*
	Kehl *et al.* full	—	—	—	—	2.553	2.644	**1.263**	**0.850**	**2.282**	**1.381**
	ours (with refin.)	**0.152**	**0.146**	**0.146**	**0.150**	**0.315**	**0.336**	1.276	1.128	2.358	1.960
leopard	KinFu	0.525	0.758	0.540	0.734	1.785	1.299	4.445	2.430	1.886	3.292
	ours (no refinement)	0.226	0.264	0.237	0.268	0.760	0.830	2.692	1.882	1.321	1.220
	Kehl *et al.* object	0.330	0.324	0.260	0.268	1.018	1.378	5.693	5.050	*failed*	*failed*
	Kehl *et al.* full	—	—	—	—	3.626	3.705	1.907	1.218	1.281	1.218
	ours (with refin.)	**0.225**	**0.263**	**0.233**	**0.266**	**0.652**	**0.614**	**1.308**	**1.154**	**1.263**	**1.111**
tank	KinFu	0.900	0.708	1.274	0.911	1.390	1.315	1.561	1.453	2.579	2.265
	ours (no refinement)	0.270	0.204	0.289	0.263	0.953	0.740	1.336	1.188	2.042	2.404
	Kehl *et al.* object	0.384	0.506	3.929	3.961	1.573	2.250	1.192	1.009	2.340	2.062
	Kehl *et al.* full	—	—	—	—	2.617	2.571	1.064	0.872	**0.946**	**0.806**
	ours (with refin.)	**0.267**	**0.199**	**0.285**	**0.263**	**0.466**	**0.416**	**0.911**	**0.745**	1.508	1.760

Conclusions

- Frame-to-frame strategy is better for object reconstruction
- Inherent smoothing properties of SDFs handle Kinect-like noise better
- SDF-2-SDF can handle smaller overlap
- E_{norm} speeds up convergence

Discussion and outlook

- Only small objects
  - Solved in follow-up with multiple limited-extent volumes
- No RGB data used
  - Energy formulation allows for additional (color) constraints
- No live reconstruction
  - Solved in follow-up with simultaneous usage of GPU and CPU
- Only static objects
  - Method used for deforming surfaces in KillingFusion

Thank you

Questions?