

SDF-TAR: Parallel Tracking and Refinement in RGB-D Data using Volumetric Registration

Miroslava Slavcheva
mira.slavcheva@tum.de

Slobodan Ilic
slobodan.ilic@siemens.com

Technische Universität München
Munich, Germany

Siemens AG
Munich, Germany

SDF-TAR is a real-time SLAM system which employs volumetric registration in RGB-D data. It is based on the SDF-2-SDF registration energy [1] that minimizes the per-voxel difference of a pair of signed distance fields (SDFs). The energy is used both in the GPU frame-to-frame tracking module, and in the concurrent CPU batch pose refinement module. To minimize runtime and memory consumption, registration is done only over several limited-extent volumes (LEVs), anchored at locations of high curvature.

LEVs The original SDF-2-SDF registration uses regular voxel grids, which become too memory-intensive if scanning large spaces with a fine resolution is desired. To tackle this issue, we carry out registration in a number of partial volumes (dubbed LEVs), which guarantee an upper bound on the runtime and memory requirements. We set them at the most geometrically discriminative regions of a scene, namely at the locations of highest curvature. These anchor points are very fast to compute as the second order derivative directly from the depth image, followed by a non-maximum suppression step. This allows us to select peaks sufficiently far apart, so that volumes do not overlap. Experiments show that this strategy leads to higher tracking accuracy than taking the same numbers of uniformly spaced or randomly placed LEVs.

Registration To estimate the camera pose ξ , represented via 6-element twist coordinates, of an SDF ϕ_c with weight field ω_c relative to a reference SDF ϕ_r with weight field ω_r generated from the identity pose, we minimize the sum of direct per-voxel differences in all p LEVs Ω_i :

$$E_{SDF}(\xi) = \frac{1}{2} \sum_{LEV \Omega_i}^{i=1..p} \left(\sum_{\text{voxels} \in \Omega_i} (\phi_r \omega_r - \phi_c(\xi) \omega_c(\xi))^2 \right). \quad (1)$$

Parallel tracking and refinement Tracking is carried out in a frame-to-frame fashion on the GPU, using a first-order Taylor approximation of

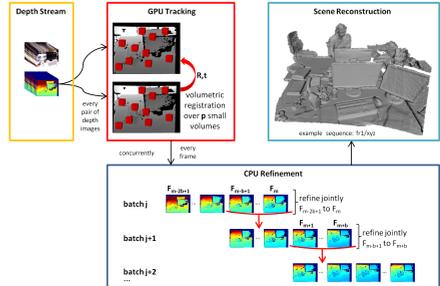


Figure 1: Overview: GPU frame-to-frame tracking, and concurrent CPU batch pose refinement.

Eq. 1 which yields a 6×6 linear system in every iteration. *Pose refinement* is done jointly over batches of $2b$ frames concurrently on the CPU, until the next batch becomes available. The second half of a batch are the most recently tracked b frames, while the first half are the previous b frames that have already been optimized once. The first $b/2$ poses are kept fixed for stability. As tracking ensures a good initialization, this optimization follows a simple gradient descent over the 6-element pose vector of every camera.

Results We evaluate tracking precision on the TUM RGB-D benchmark [2]. Via the use of LEVs, SDF-TAR disregards regions that could impede registration, leading to considerably better rotational and on-par translational motion estimation with related volumetric techniques. In addition, we achieve higher reconstruction fidelity on the 4 objects of the CoRBS dataset [3].

- [1] M. Slavcheva, W. Kehl, N. Navab, and S. Ilic. SDF-2-SDF: Highly Accurate 3D Object Reconstruction. In *Proc. ECCV*, 2016.
- [2] J. Sturm, N. Engelhard, F. Endres, W. Burgard, and D. Cremers. A Benchmark for the Evaluation of RGB-D SLAM Systems. In *Proc. IROS*, 2012.
- [3] O. Wasenmüller, M. Meyer, and D. Stricker. CoRBS: Comprehensive RGB-D Benchmark for SLAM using Kinect v2. In *Proc. WACV*, 2016.