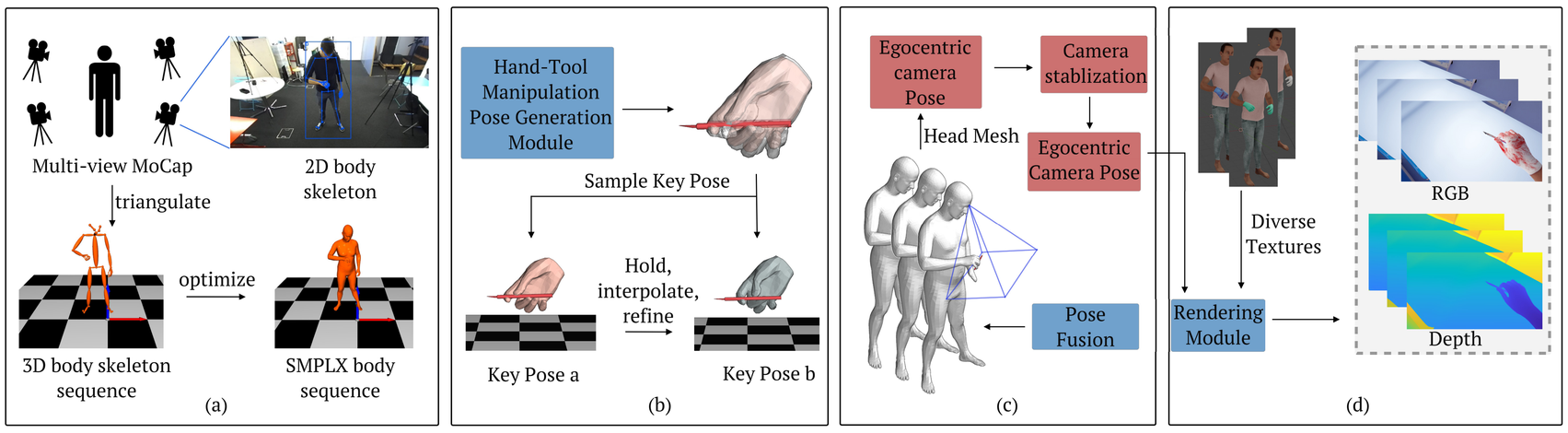
POV-Surgery: A Dataset for Egocentric Hand and Tool Pose Estimation During Surgical Activities

Rui Wang, Sophokles Ktistakis, Siwei Zhang, Mirko Meboldt and Quentin Lohmeyer

We propose a novel synthetic data generation pipeline that produces hand-tool manipulation temporal sequences from egocentric perpective. Using the data generation pipeline and focusing on three tools used in orthopedic surgeries: scalpel, diskplacer, and friem, we propose a large, synthetic, and temporal dataset on egocentric surgical hand-object pose estimation.We fine-tune the current SOTA methods on POV-Surgery and further show the generalizability when applying to real-life cases with surgical gloves and tools by extensive evaluations.

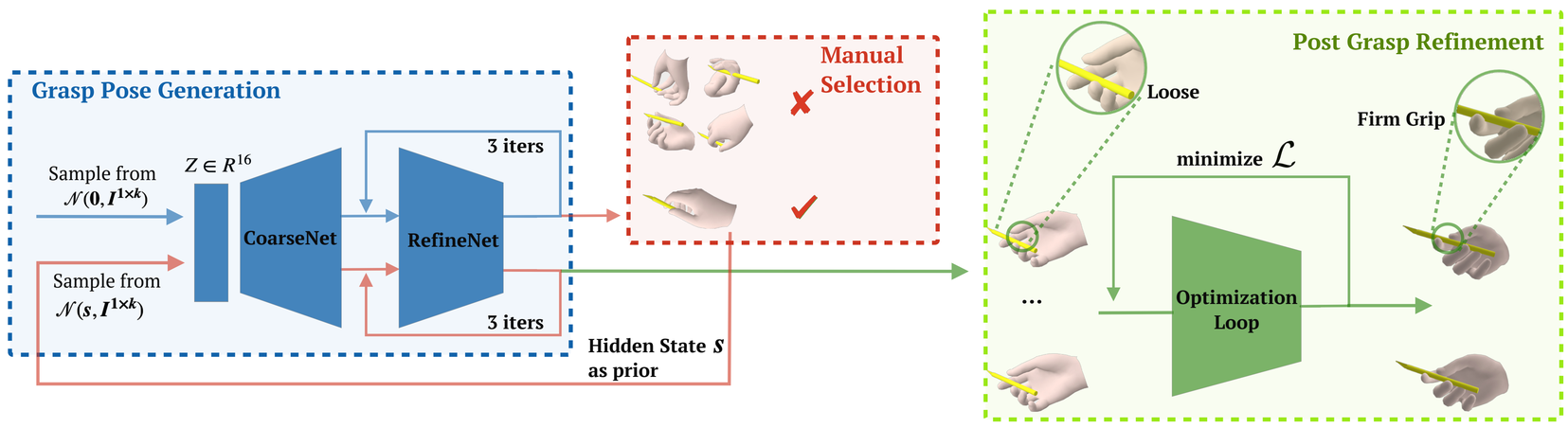


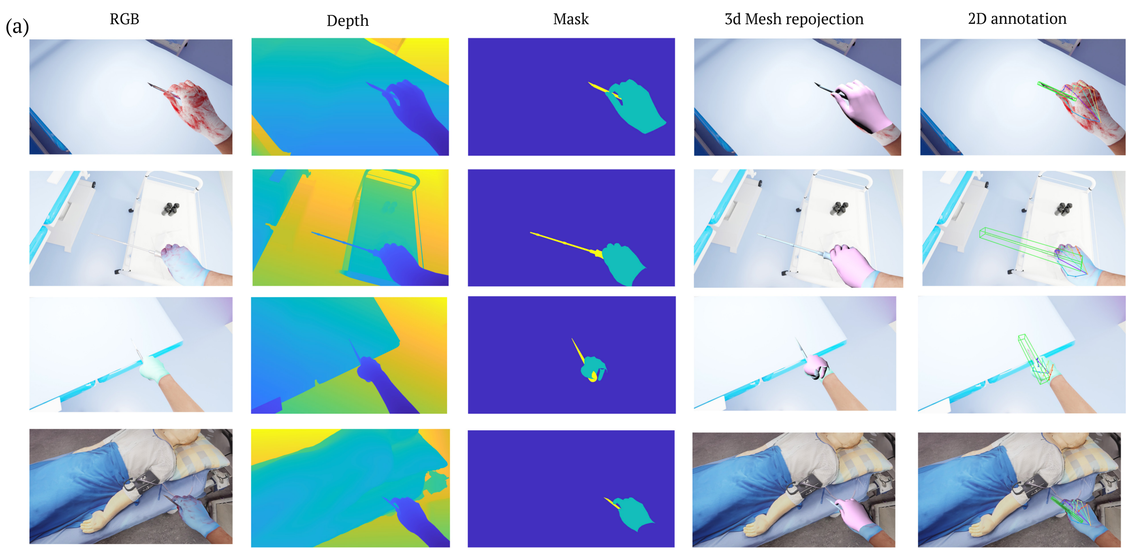
Data Generation Pipeline

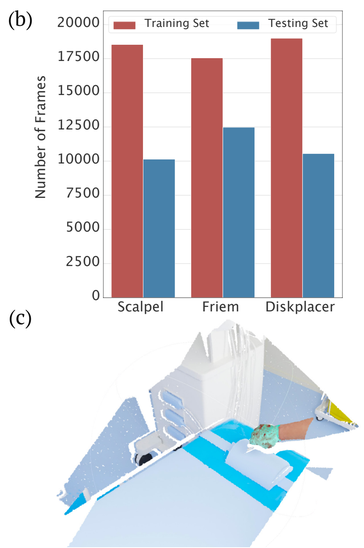
The proposed pipeline to generate synthetic data sequences. **(a)** shows the multi-stereo-cameras-based body motion capture module. **(b)** shows the optimization- based hand-object manipulation sequence generation pipeline. **(c)** shows the fused hand-body pose and the egocentric camera pose calculation module. **(d)** shows the rendering module with which RGB-D sequences are rendered with diverse textures.

Grasp Generation

The Hand manipulation sequence generation pipeline consists of three components: grasp pose generation, pose selection, and pose refinement, highlighted in blue, red, and green, respectively.







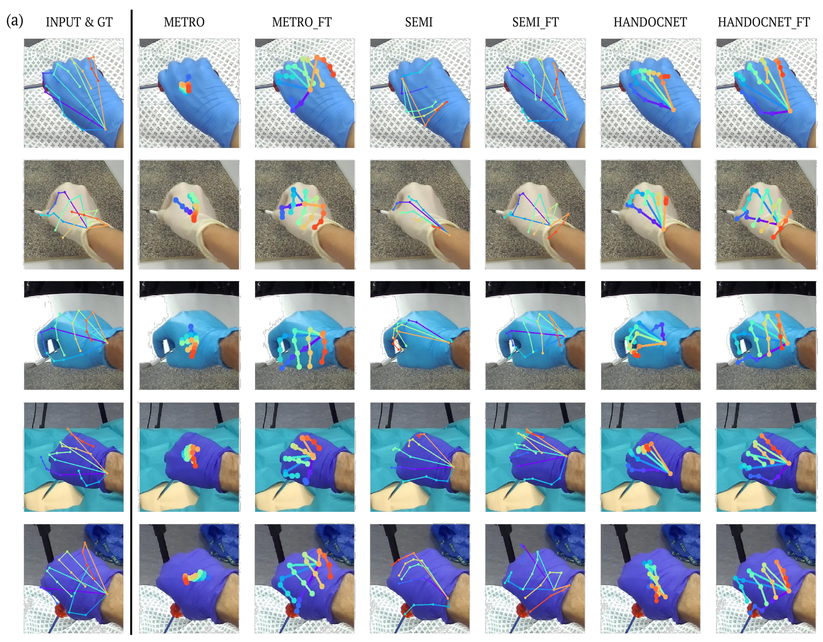
Data Overview

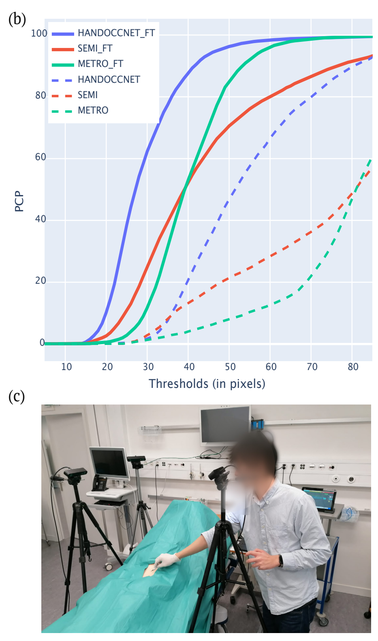
(a) Dataset samples for RGB-D sequences and annotation. An example of the scalpel, friem, and diskplacer, is shown in the first three rows. The fourth row shows an example of the new scene and blood glove patterns that only appear in the test set. **(b)** shows the statistics on the number of frames for each surgical instrument in the training and testing sets. **(c)** shows a point cloud created from an RGB-D frame with simulated Kinect noise.

Benchmark

The evaluation result of different SOTA methods on the test set of POV-Surgery dataset.

Method	P_{2d} ↓	MPJPE ↓	PVE ↓	PA-MPJPE ↓	PA-PVE ↓
METRO	95.11	77.46	75.06	23.43	22.34
SEMI	77.91	115.67	112.10	12.68	12.76
HandOCCNet	64.70	95.19	90.83	11.71	11.13
METRO$_{ft}$	30.49	14.90	13.80	6.36	4.34
SEMI$_{ft}$	**13.42**	15.14	14.69	**4.29**	**4.23**
HandOCCNet$_{ft}$	13.80	**14.35**	**13.73**	4.49	4.35





Real-life Evaluation

(a) Ground truth and qualitative results of different methods on the real-life test set. **(b)** Accuracy with different 2D pixel error thresholds, showing large performance improvement after fine-tuning on POV-Surgery **(c)** Our multi-camera real-life data capturing set-up.

Scan to see project page for details!