A Dataset of Human Knot Tying Demonstrations with Paired Rope Topology and Hand Trajectories

Youngbin Yoon[1], Ayoung Hong[1]

[1]Department of Mechanical Engineering, Chonnam National University

Introduction

- Knot tying is a fundamental yet challenging deformable object manipulation task, requiring the handling of complex rope deformations, frequent self-occlusions, and ambiguous crossings that make perception and control difficult.
- Most Learning-from-Demonstration (LfD) studies focus on reproducing human hand trajectories [1], which is insufficient for DLOs, as identical motions can result in different shapes.
- What remains consistent across demonstrations is the sequence of topological states, suggesting that learning should focus on how actions transform topology rather than merely replicating motion.
- We present a stage-structured human demonstration dataset that links hand trajectories with rope topology to enable studies on topology-aware knot tying.

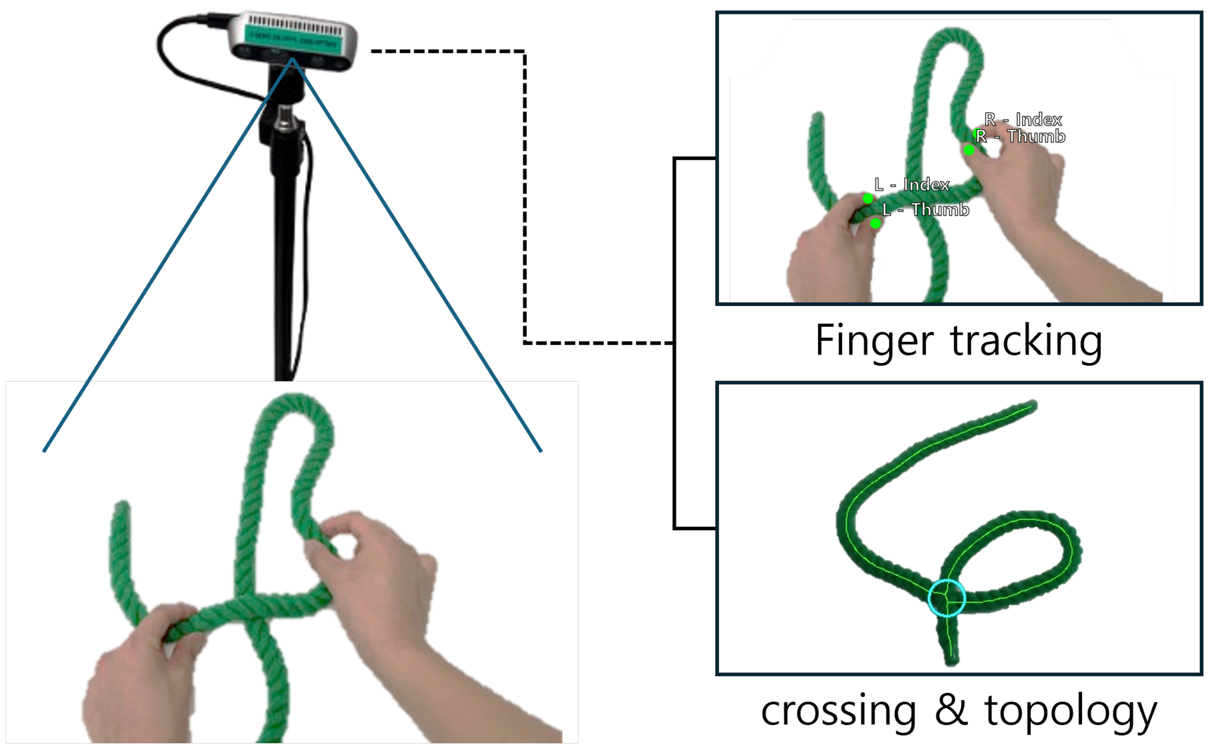

Figure 1: Experiment environment

Dataset: Human Knot-tying Demonstrations

- 32 sessions (5 participants $\times$ 2 knot types $\times$ 3 trials + ideal cases) recorded with an Intel RealSense D435 (RGB-D @ 30 Hz).
- Stages defined when both hands left the view for $\geq 1\,\mathrm{s}$ (τ_{gap}).
- Each stage includes 3D fingertip keypoints, rope topology from rope-only frames, and synchronized RGB-D images.
- Representative frame sampled $0.5\,\mathrm{s}$ before the next hand-in to ensure only the rope is visible.

Dataset: Statistics & Variability

- For every stage, the following data are stored:
  - RGB-D frames of the corresponding segment
  - Rope topology annotations: the detected centerline of the rope and the number of crossings
  - 3D hand keypoints (thumb and index of both hands)
  - A representative frame (`rep_color`, `rep_topology`) showing the rope before manipulation
- Participants tied knots in their own preferred way, without fixed grasping orders or motions.
- As a result, tying sequences, stage durations, and grasping positions differ across sessions, while the stage-level rope topologies remain consistent.
- This structure allows analysis of how different human strategies produce the same topological transitions.

Knot Type	Session Type	Sessions	Avg.Stages	Avg.Crossings per Stage	Avg.Final Crossing
Overhand	Ideal	1	3	0 / 1 / 2 / .	3
Overhand	Human	15	3.13	0 / 1 / 2 / 2.5	2.07
Figure-Eight	Ideal	1	4	0 / 1 / 2 / 3	4
Figure-Eight	Human	15	3.6	0.13 / 1.33 / 1.86 / 2.5	0.93

Table 1: Dataset statistics by knot type and session type.

Acknowledgement: This work was supported in part by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (RS-2024-00340040) and in part by Korea Institute for Advancement of Technology(KIAT) grant funded by the Korea Government(MOTIE)(RS-2024-00406796, HRD Program for Industrial Innovation)

Topology Analysis

- Stage-level rope topologies were visualized to track how crossings change during knot tying. Detected clusters indicate where rope strands intersect, forming consistent topological snapshots across stages and participants.

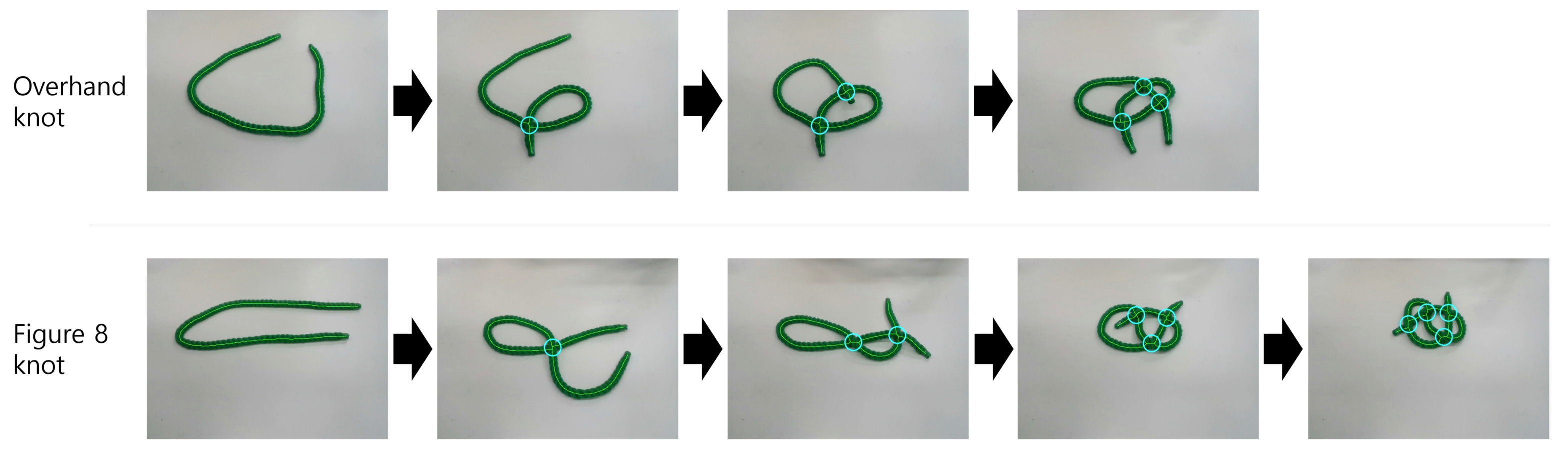

Figure 2: Stage sequences for Overhand and Figure-Eight knots.

- Rope topology was computed from rope-only frames using the **8-neighborhood count-group rule**. A pixel is identified as a crossing when multiple distinct neighbor groups surround it as follows:

$$N \leq 2 : \text{non-crossing}$$
$$N \in \{3,4\} \text{ and } G \geq 3 : \text{crossing}$$
$$N \geq 5 : \text{crossing}$$

where N is the number of occupied neighbor pixels, and G is the number of 4-connected groups (diagonals excluded).

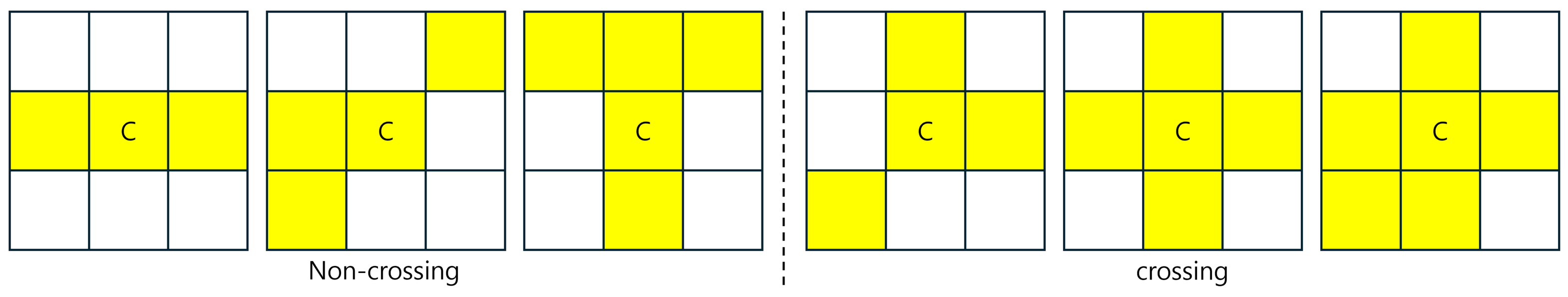

Figure 3: 8-neighborhood rule for crossing detection.

- Depth information distinguishes upper and lower strands, producing reliable crossing maps for each stage.

Limitations and Challenges

- This dataset currently has several constraints as summarized below:
  - Crossing detection fails when rope segments are tightly overlapped or twisted, as self-occlusion prevents crossings from being recognized.
  - The 8-neighborhood rule may misclassify pixels when nearby strands belong to the same contact region.
  - Depth resolution is often insufficient to determine over/under relationships when strand separation is smaller than sensor precision.
  - The dataset currently covers only two knot types with a small number of participants, limiting generalization.

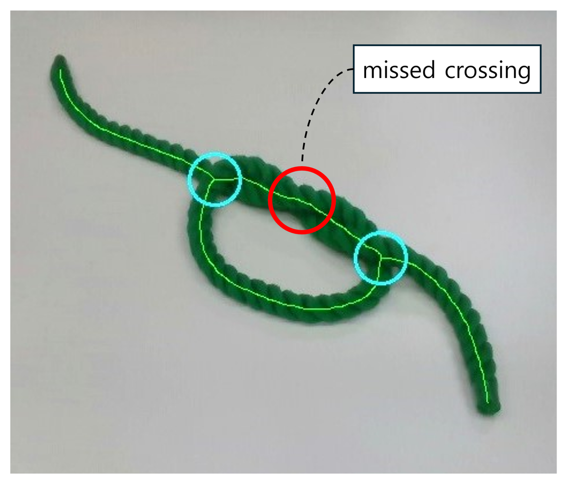

Figure 4: Failure case

Conclusion and Future Work

- We presented a stage-structured dataset of human knot-tying demonstrations, aligning rope-only snapshots with fingertip trajectories to analyze manipulation as a sequence of topological transitions.
- Current limitations include overlap-induced detection errors, depth ambiguity, and limited diversity of knot types and participants, which will be addressed in future releases with additional data and improved perception.
- Future work will focus on understanding rope topology and its relation to crossing counts and tying strategies. We aim to develop topology-aware policy learning that manipulates the rope from its current topology to form a desired knot.

[1] Ravichandar, Harish, et al. "Recent advances in robot learning from demonstration." Annual review of control, robotics, and autonomous systems 3.1, 2020.

[2] Schulman, John, et al. "Learning from demonstrations through the use of non-rigid registration." Robotics Research: The 16th International Symposium ISRR. Cham: Springer International Publishing, 2016.

[3] Dinkel, Holly, et al. "KnotDLO: Toward Interpretable Knot Tying." arXiv preprint arXiv:2506.22176 (2025).