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

Youngbin Yoon and Ayoung Hong*
Department of Mechanical Engineering, Chonnam National University

***Abstract*—Learning from demonstration (LfD) for deformable linear objects such as ropes requires more than imitating hand trajectories. Grasping positions, substep order, and motions vary across demonstrators, yet rope topology at stage boundaries converges to consistent states. To study knot tying conditioned on topology, we present a small but structured dataset of 30 human demonstrations of the Overhand and Figure-Eight knots, tied by five participants. Each session is segmented into stages by brief hand removal, yielding rope-only frames for clean topology snapshots. For every stage, we provide (i) 3D fingertip keypoints, (ii) rope crossings with depth-based over/under labels, and (iii) synchronized RGB-D imagery. We also include demonstrations with separated strands to provide clean cases, while real tying often produces overlaps and self-occlusions that challenge current algorithms. Rather than a large benchmark, this dataset offers a starting point for studying knot tying as a sequence of topology transitions and for developing policies that go beyond trajectory imitation.**

I. Introduction

Deformable linear objects (DLOs) such as ropes, cables, and sutures appear in many domains including surgery, household robotics, and industrial automation. Knot tying is an important operation in these settings, yet it poses significant challenges. Unlike rigid bodies, DLOs do not preserve a fixed shape: small differences in applied forces or grasp points can lead to different configurations, making actions difficult to reproduce. Perception is also difficult because ropes often overlap with themselves, producing self-occlusions that hide crossings. Thus, geometry alone is insufficient to describe the rope's state, and reasoning about topological changes across the tying process is essential.

Most prior work in robotic learning from demonstration (LfD) has focused on imitating hand trajectories [1]. This is inadequate for knot tying: people vary in grasping positions, substep order, and hand motions, and even the same person cannot reproduce identical trajectories because the rope deforms differently each time. Instead, what remains consistent across demonstrations is the sequence of topological states, suggesting that policies should focus on how actions transform topology from one stage to the next.

To support this direction, we present a human–rope dataset for knot tying explicitly structured into stages. Each stage boundary is marked by both hands leaving the camera view for about one second, producing rope-only frames for clean topology snapshots. The dataset covers two canonical knots—the Overhand and the Figure-Eight—tied by five participants with three repetitions each, yielding 30 sessions in total. For every stage, we provide (i) 3D fingertip keypoints, (ii) rope topology represented as clustered crossings, and (iii) synchronized RGB-D imagery. By organizing demonstrations this way, the dataset links human manipulation with topological transitions and offers a resource for policies that go beyond trajectory imitation.

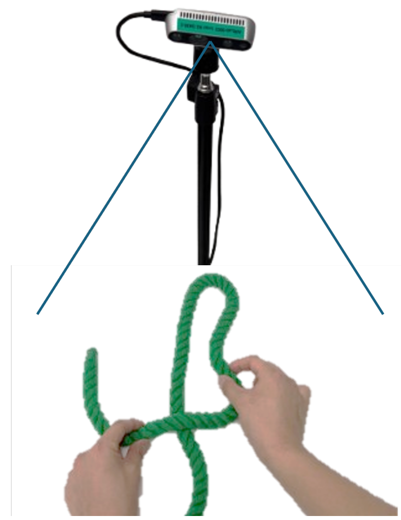

Fig. 1. Experimental setup for data collection. Demonstrations were recorded with an Intel RealSense D435 mounted in front of the workspace. Participants tied knots on a tabletop area within the camera's field of view, following the stage protocol where hands are briefly removed to segment substeps.

II. Related Work

Research on learning from demonstration (LfD) has explored many ways to transfer human demonstrations to robots. Most work has focused on trajectory reproduction, but its limitations have motivated recent studies on learning action policies instead [1]. For non-rigid objects, non-rigid registration has been applied to align demonstrations with new conditions [2], but these methods only generalized the initial topology and did not address the state changes central to knot tying.

Perception-driven approaches have been studied mainly in the context of rope disentangling. Some methods learn task-relevant keypoints to represent rope state [3], while others combine RGB-D input with a particle filter to maintain a distribution over possible configurations [4], or treat multi-cable disentangling as a graph problem [5]. These works

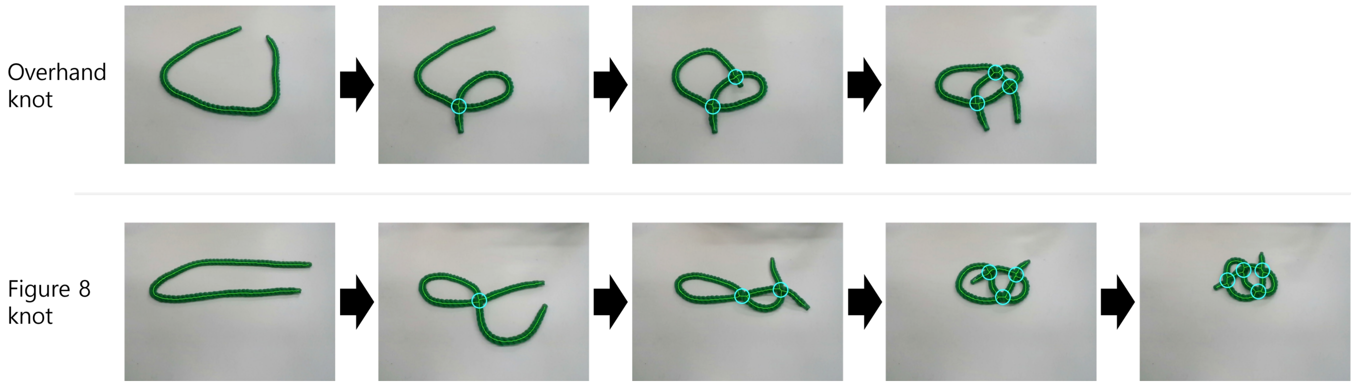

Fig. 2. Stage sequences for Overhand and Figure-Eight knots. Each stage shows rope-only frames with detected crossing clusters (cyan circles). While participants vary in order and grasp sites, stage-level topologies converge to consistent intermediate states.

highlight the importance of perception for entangled structures, but focus on untangling rather than tying.

Another line of work investigates rope tracking under occlusion. TrackDLO introduced motion-coherent tracking [6], and geometric constraints have been used to enforce consistency [7]. While these approaches improve robustness, they still fail when ropes are tightly overlapped or tied into knots, where the shape becomes highly ambiguous. Similarly, manipulation skills learned from dense object descriptors trained on synthetic depth data [8] remain trajectory-centered and do not reason about topology.

Explicitly topological approaches are rarer. KnotDLO [9] used topological waypoints to enable one-handed overhand knot tying, showing that reasoning at the level of topology supports repeatable tying. However, the approach was validated only on a single knot—the Overhand knot—limiting its generality to other types of knots.

III. Dataset Design

We designed our dataset to capture human demonstrations of knot tying with explicit stage segmentation and paired annotations of hand trajectories and rope topology. Demonstrations were recorded with an Intel RealSense D435 camera, logging RGB images at 1280×720 and depth images at 848×480, both at 30 Hz. Each session consists of a continuous recording of a complete knot tying attempt. To segment the process into stages, participants were instructed to remove both hands from the camera view after completing each substep. When both hands were absent for at least $\tau_{\text{gap}} = 1.0\,\text{s}$, the current stage was closed, producing rope-only frames that serve as clean snapshots of rope topology. For the next stage, a representative frame was chosen 0.5 s before the hands reappeared, ensuring that the snapshot contained no visible hand pixels.

The dataset covers two canonical knots: the Overhand and the Figure-Eight. Five participants tied each knot three times, yielding a total of 30 recorded sessions. This protocol captures natural variation across participants and repetitions, while converging to consistent topological states at stage boundaries.

Annotations include both hand and rope information. Hand trajectories are obtained by detecting the thumb and index fingertips in each frame and projecting them into 3D using synchronized depth images. Rope topology is derived from rope-only frames at stage boundaries. After HSV-based segmentation and skeletonization, crossings are identified using an 8-neighborhood count–group rule. For each skeleton pixel C, we define N as the number of occupied neighbor pixels among the 8 surrounding locations, and G as the number of contiguous groups of such pixels when checked under 4-connectivity (up, down, left, right). A pixel is classified as non-crossing if $N \leq 2$; for $N \in 3, 4$, it is a crossing only when $G \geq 3$; and for $N \geq 5$, it is always considered a crossing. This criterion eliminates spurious detections caused by diagonal contacts or curved T-junctions, while robustly capturing genuine intersections (see Figure 3). Detected crossings are clustered into nodes, and synchronized depth information is used to disambiguate the upper and lower strands at each intersection.

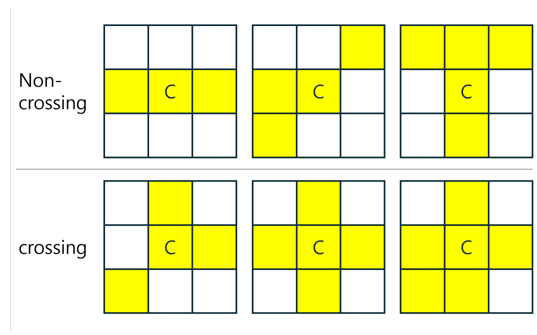

Fig. 3. Example of the 8-neighborhood rule for crossing detection. The center pixel is evaluated based on the number and grouping of occupied neighbors to distinguish true crossings from simple contacts.

For every stage, we provide (i) fingertip trajectories (`hands.jsonl`), (ii) synchronized RGB-D frames, and (iii) representative rope-only snapshots (`rep_color`, `rep_topology`) together with annotated crossings. At the session level, a global `timeline.json` indexes all stages, links them to the raw recordings, and stores the final topology at the end of the sequence.

Fig. 4. Hand keypoint tracking during tying. Thumb and index fingertips are annotated in 3D and projected onto RGB-D frames, linking grasping actions with rope topology changes.

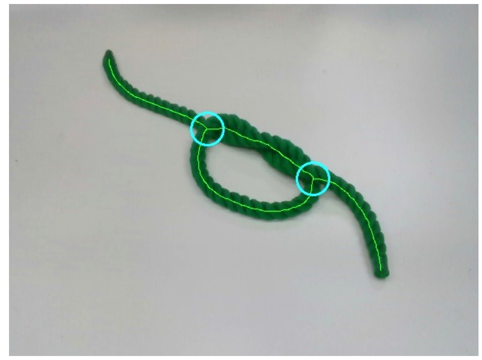

Fig. 5. Failure case where rope segments overlap tightly. Although the rope is twisted, the algorithm fails to detect a crossing due to overlap and self-occlusion, illustrating a key challenge for robust topology perception.

IV. Analysis and Challenges

At the stage level, our dataset provides clean and interpretable supervision. Figure 2 illustrates the stage sequences for the Overhand and Figure-Eight knots. Each stage is associated with a rope-only snapshot and annotated crossings. Although participants vary in grasping positions and hand motions, the rope topology at stage boundaries converges to consistent intermediate structures. This demonstrates that stage-based segmentation successfully aligns human demonstrations with repeatable topological states, offering a reliable foundation for policy learning.

Hand information is aligned with rope topology to link manipulation with topological changes. Figure 4 shows an example of fingertip tracking during the tying process. The thumb and index fingertips are detected in 3D and projected onto the RGB-D frames, illustrating how participants interact with rope segments. These annotations enable analysis of how human tying actions are directly connected to topological configurations defined by crossings.

At the same time, the dataset reveals clear limitations of our crossing detection method. Figure 5 shows a failure case where rope segments are tightly overlapped or twisted. Although crossings are physically present, visual overlap and self-occlusion prevent them from being recognized. In particular, our 8-neighborhood rule can misclassify pixels when neighboring strands lie very close together: N may exceed the threshold, but the strands belong to the same physical contact rather than a true crossing. Moreover, depth information is often insufficient when the separation between strands is below the sensor's resolution, making over/under assignment unreliable. Such cases occur frequently in natural tying and remain a key challenge for robust topology-aware perception.

V. Conclusion

In this work, we introduced a stage-structured dataset of human knot tying demonstrations with paired annotations of hand trajectories and rope topology. The dataset is designed to study knot tying not only as a problem of reproducing human motion, but as a sequence of topological transitions. By aligning rope-only snapshots with hand keypoints at each stage, the dataset provides clean supervision for investigating how manipulation actions transform rope topology.

At the same time, the dataset has clear limitations. It currently covers only two knots, the Overhand and the Figure-Eight, with a small number of participants. Although depth information is used to distinguish upper and lower strands at crossings, failures still occur when rope segments are tightly overlapped or heavily self-occluded. In future work, we plan to extend the dataset to include additional knot types, more participants, and multi-view recordings, as well as richer annotations such as explicit over/under depth labeling.

We hope this dataset will serve as a useful resource for the community, not as a large-scale benchmark but as a step toward studying knot tying as a sequence of topology transitions. We also welcome feedback on how to improve the dataset design and annotations, and look forward to collaboration on developing more robust topology-aware policies for deformable object manipulation.

References

[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] Grannen, Jennifer, et al. "Untangling dense knots by learning task-relevant keypoints." arXiv preprint arXiv:2011.04999, 2020.
[4] Lui, Wen Hao, and Ashutosh Saxena. "Tangled: Learning to untangle ropes with rgb-d perception." 2013 IEEE/RSJ International Conference on Intelligent Robots and Systems, 2013.
[5] Viswanath, Vainavi, et al. "Disentangling dense multi-cable knots." 2021 IEEE/RSJ International Conference on Intelligent Robots and Systems, 2021.
[6] Xiang, Jingyi, et al. "Trackdlo: Tracking deformable linear objects under occlusion with motion coherence." IEEE Robotics and Automation Letters 8.10, 2023
[7] Wang, Yixuan, Dale McConachie, and Dmitry Berenson. "Tracking partially-occluded deformable objects while enforcing geometric constraints." 2021 IEEE International Conference on Robotics and Automation, 2021.
[8] Sundaresan, Priya, et al. "Learning rope manipulation policies using dense object descriptors trained on synthetic depth data." 2020 IEEE International Conference on Robotics and Automation, 2020.
[9] Dinkel, Holly, et al. "KnotDLO: Toward Interpretable Knot Tying." arXiv preprint arXiv:2506.22176 (2025).