Autonomous Needle Manipulation for Robotic Surgical Suturing Based on Skills Learned from Demonstration

Schwaner, Kim Lindberg; Dall'Alba, Diego; Jensen, Pernille Tine; Fiorini, Paolo; Savarimuthu, Thiusius Rajeeth

Published in:
2021 IEEE 17th International Conference on Automation Science and Engineering (CASE)

DOI:
10.1109/CASE49439.2021.9551569

Publication date:
2021

Document version:
Accepted manuscript

Citation for pulished version (APA):
Schwaner, K. L., Dall'Alba, D., Jensen, P. T., Fiorini, P., & Savarimuthu, T. R. (2021). Autonomous Needle Manipulation for Robotic Surgical Suturing Based on Skills Learned from Demonstration. In *2021 IEEE 17th International Conference on Automation Science and Engineering (CASE)* (pp. 235-241). IEEE. IEEE International Conference on Automation Science and Engineering Vol. 2021-August
https://doi.org/10.1109/CASE49439.2021.9551569

Autonomous Needle Manipulation for Robotic Surgical Suturing Based on Skills Learned from Demonstration

Kim Lindberg Schwaner* Diego Dall'Alba† Pernille Tine Jensen‡ Paolo Fiorini†
Thiusius Rajeeth Savarimuthu*

Abstract

In the future, surgical robots will grant the option of executing surgical tasks autonomously, supervised by the surgeon. We propose a simple framework for learning surgical action primitives that can be used as building blocks for composing more elaborate surgical tasks. Our method is based on Learning from Demonstration (LfD) as this allows us to exploit existing expert knowledge from recordings of surgical procedures. We demonstrate that we can learn needle manipulation actions from human demonstration, constructing an *action library* which is used to autonomously execute part of a surgical suturing task. Actions are learned from single demonstrations and we use Dynamic Movement Primitives (DMPs) to encode low-level Cartesian space trajectories. Our method is experimentally validated in a non-clinical setting, where we show that learned actions can be generalized to previously unseen conditions. Experiments show a 81 % task success rate for moderate variations from the initial conditions of the demonstration with a mean needle insertion error of 3.8 mm.

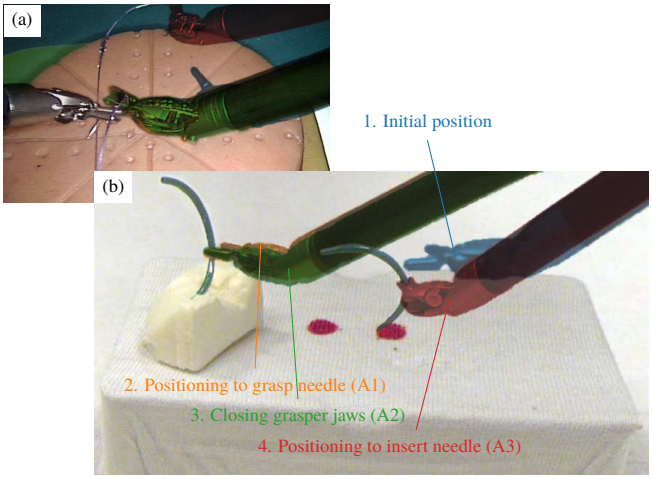


Figure 1: Automating part of a surgical suturing task: (a) Demonstrated gestures used to build a library of surgical actions; (b) Autonomously performed sequence of actions to grasp and position the needle under previously unseen task conditions.

1 Introduction

Robot-assisted Minimally Invasive Surgery (RMIS) facilitates a growing number of surgical procedures with positive impact on patient outcomes [1]. Similar to what we see in industry, we expect that robots will enable autonomous execution of surgery, under supervision by the surgeon [2]. This has the potential to make surgery more consistent and shorten waiting time in hospitals. A surgical procedure involves a number of phases which can be decomposed into smaller tasks like dissecting or suturing. Designing specific controllers for each task is challenging and time-consuming, since it is necessary to determine optimal parameters by hand for a large number of tasks. We instead envision the surgical robot system as an apprentice to the surgeon, i.e., it automatically learns while it is being operated by the surgeon to ultimately become able to autonomously perform learned tasks.

In this paper we present early results toward using expert demonstrations for automating surgical tasks with a novel application of Learning from Demonstration (LfD) techniques in RMIS. We propose a framework for building an *action library* of surgical action primitives in a manner that is independent of specific robot systems and generalizes to new and different task conditions. We show that we can learn actions from human demonstration, bridging the gap between surgeon and automated robot system. The proposed framework is applied to learn a number of needle manipulation actions which are then arranged, as shown in Fig. 1, to form part of a surgical suturing task. The task is subsequently executed in a different context, compared to the original demonstration. Our approach is evaluated on a da Vinci Research Kit (dVRK) [3] and incorporates a stereo computer vision system to perceive the environment and a task planner module to adapt actions to it.

Previous studies have investigated automation of surgical tasks or part thereof. The work in [4]–[13] focuses on suturing needle grasping and insertion as well as optimal selection of needle shape, size and grasp. Other studies have dealt with automating suturing using custom instruments [14], [15] and yet others with tasks like cutting [16], [17], knot tying [18]–

*K. L. Schwaner and T. R. Savarimuthu are with SDU Robotics, The Mærsk Mc-Kinney Møller Institute, Faculty of Engineering, University of Southern Denmark, 5230 Odense, Denmark; {kils,trs}@mmmi.sdu.dk.

†D. Dall'Alba and P. Fiorini are with the Department of Computer Science, University of Verona, 37134 Verona, Italy; {diego.dallalba,paolo.fiorini}@univr.it.

‡P. T. Jensen is with the Department of Gynaecology and Obstetrics, Aarhus University Hospital, 8200 Aarhus, Denmark; petije@rm.dk.

[20] and the removal of phantom tissue [21]–[23]. Many of the listed studies have relied on purely computer generated motion for performing tasks autonomously [6]–[13], [22], [23]. In this work we take advantage of the fact that systems for RMIS are typically tele-operated and so invite LfD. Other studies have also taken an LfD approach [4], [5], [16], [18]–[21].

A notable limitation of some of the methods proposed in the literature is the inability to generalize to differing task conditions [16], [18] or only to task conditions close to those of the initial demonstrations [19], [20]. Another frequently seen limitation is requiring human input to perform one or more critical steps of the task, e.g., registering tasks to the environment [4], [5], suture needle insertion [4] and grasping [4], [13], [18], [20], [21].

One important issue in connection with autonomous suturing is the localization and manipulation of the needle. Some researchers used a mechanical guide, mounted on the grasper jaws of a surgical instrument, to artificially constrain the grasp pose of the suturing needle [10], [13]. While effective, this modification of the surgical instrument makes it unsuitable for clinical use. In [14], [15], the issue of grasping the suturing needle was completely avoided by using a special instrument for placing sutures.

The method we propose is based on LfD, but stands out from previous work in that we use a consolidated library of primitive actions learned from demonstration. We further demonstrate using the action library for entirely LfD-based autonomous manipulation of a suturing needle, including grasping, under previously unseen task conditions and using unmodified surgical instruments.

The remainder of this paper is organized as follows: Section 2 describes our framework for learning and executing surgical actions. In Section 3 we detail a needle manipulation task that is executed autonomously under previously unseen task conditions. Section 4 reports our experimental results, Section 5 brings a discussion of the results and future work and Section 6 concludes the paper.

2 Surgical Action Framework

An overview of the proposed learning framework is given in Fig. 2 and shows two main phases: learning and execution. The learning phase begins with recordings of a human operator performing a given task. The manipulator kinematics and external sensor time series data are saved in a database. Surgical motions are manually segmented from the time series and used to learn action descriptors, which are subsequently stored in an *action library*. Action descriptors store motions, encoded using Dynamic Movement Primitives (DMPs), and their relation to specific landmarks. This enables us to generalize actions to new task contexts, i.e., the geometric conditions of the task, in the execution phase.

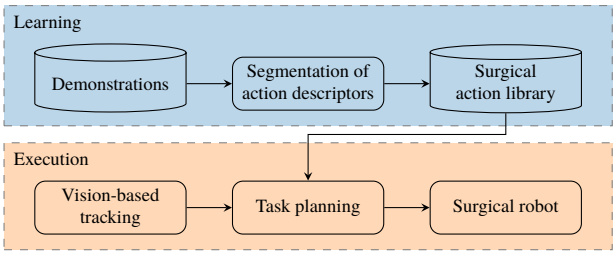


Figure 2: The workflow of the framework. In the learning phase, human demonstrations are segmented and used for building action descriptors. In the execution phase, actions are used to realize a given surgical task.

Field description	Symbol(s)
Basis function weights	$\{\boldsymbol{w}^p, \boldsymbol{w}^o, \boldsymbol{w}^g\}$
Goal position	$\{\boldsymbol{g}^p, \boldsymbol{g}^o, \boldsymbol{g}^g\}$
Duration	τ
Target object label	$\mathcal{O}$
Reference frame	$\mathcal{F}$

Table 1: Surgical action descriptor.

2.1 Surgical Actions

We define surgical actions as primitive, non-overlapping sequences of a surgical task. Actions describe the kinematic state of a single manipulator throughout its duration and a geometric relation to an object or landmark in the environment. An action can be generalized to a new task context by registering it to observed objects or landmarks. Table 1 lists an *action descriptor*, which constitutes a single entry in the surgical action library.

We choose to encode the kinematic state of a manipulator using linear DMPs [24], [25]. Here we use the Cartesian space DMP formulation by [26], encoding orientations with unit quaternions. We represent the manipulator gripper closing angle with an additional one-dimensional DMP. Thus, the full kinematic trajectory (position, orientation and gripper opening angle) of the manipulator is encoded in a set of basis function weights $\{\boldsymbol{w}^p, \boldsymbol{w}^o, \boldsymbol{w}^g\}$ and goal states $\{\boldsymbol{g}^p, \boldsymbol{g}^o, \boldsymbol{g}^g\}$. The DMPs of the positional, orientation and gripper angles are synchronized by a common *phase* [26]. We set τ equal to the duration of the demonstrated motion. The target object label, $\mathcal{O}$, identifies an object in the environment and the reference frame, $\mathcal{F}$, is a coordinate frame rigidly attached to it.

2.2 Learning from Human Demonstrations

To learn surgical actions from human demonstration we need to learn the parameters that encode the manipulator trajectory, the target object, $\mathcal{O}$, and reference frame, $\mathcal{F}$.

Given a demonstration of a surgical action, the first step in building an action descriptor is to determine what the demonstrated motion is carried out relative to. In this work, we make the assumption that the motion of an action is always made to either manipulate an object *or* position the manipulator to subsequently manipulate an object. Thus, we transform the

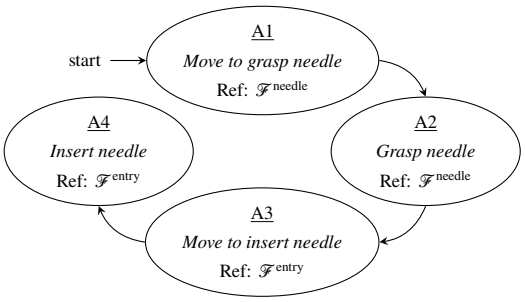


Figure 3: Finite State Machine (FSM) representing a simple surgical needle manipulation task.

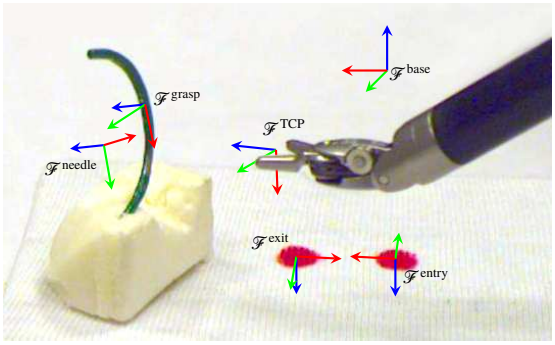


Figure 4: Context (with rigidly attached coordinate frames) of the surgical needle manipulation task.

demonstrated motion into the local coordinate frame of the target object to be manipulated. This allows motion to be expressed independently of both the robot and the environment as a whole. Once we have the demonstrated motion in relation to the target object, we find the set of basis function weights $\{\boldsymbol{w}^p, \boldsymbol{w}^o, \boldsymbol{w}^g\}$ by least-squares solutions to the systems of linear equations in [26]. The set of goal states, $\{\boldsymbol{g}^p, \boldsymbol{g}^o, \boldsymbol{g}^g\}$, is similarly expressed relative to $\mathcal{F}$.

2.3 Generalizing Actions to New Task Contexts

When adapting actions to be executed in a new task context, they are instantiated from the action library with new task-specific target object and reference frame. Thus, to apply an action in a new task context, we need to locate a suitable target object and estimate a reference frame based on sensor feedback. Additionally, the target object and reference frame need to be tracked continually, so that the motion of the action can be adapted on-line.

3 Autonomous Needle Manipulation Task

We demonstrate the proposed surgical action framework by applying it to a needle manipulation task for surgical suturing. To realize this task, we first construct an action library consisting of four surgical actions:

A1 move to grasp needle;

A2 grasp the needle;

A3 position the needle to insert it; and

A4 push the needle in.

Each action is learned from human demonstration.

The needle manipulation task is modeled as a simple Finite State Machine (FSM), diagrammed in Fig. 3, where each state instantiates an action from the action library. The actions are registered to the task execution context as it is perceived at the time the FSM transitions state, based on input from a vision-based tracking system.

3.1 Constructing an Action Library from Demonstrations

Demonstration data was sourced from a dataset consisting of 20 trials of the same four-throw suturing task performed by a single experienced user (although not a surgeon) on a first-generation da Vinci surgical robot with a dVRK. For each trial, kinematics from all Patient Side Manipulators (PSMs) and endoscope stereo video frames were recorded and manually annotated following an extended JIGSAWS convention [27].

Manipulator time-series were extracted from the demonstration dataset by first choosing a single one of the 20 trials. From that trial we find one of each of the gestures annotated

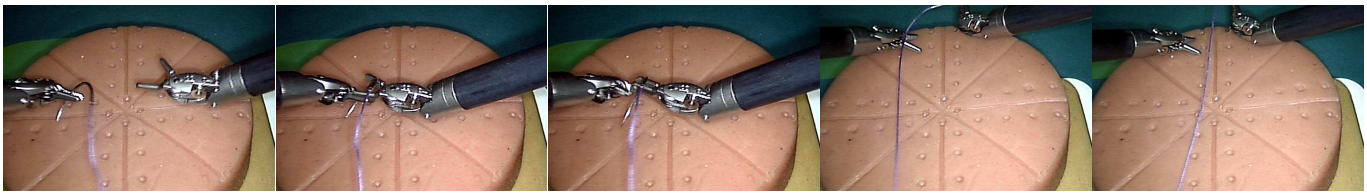
- "reaching for the needle with the right hand (G1)";
- "positioning needle [with right hand] (G2)"; and
- "pushing the needle through tissue [with the right hand] (G3)".

Gesture G2 is further divided into a "positioning to grasp needle" gesture and a "grasping needle" gesture to allow for more flexibility in the action library. Those two gestures are used to learn actions A1 and A2. The gestures G2 and G3 are used for learning actions A3 and A4 directly.

Since the calibration between the PSM1 base frame and the endoscope, ${}^{\text{base}}_{\text{endoscope}}\boldsymbol{T}$, was not included in the dataset, it was estimated after the fact by extracting corresponding PSM1 Tool Center Point (TCP) 3D points, given w.r.t. the PSM1 base frame, $\mathcal{F}^{\text{base}}$, and their 2D image projections, then solving the perspective-n-point problem.

Fig. 4 shows the coordinate frames involved in the task, providing task context. Because actions A1 and A2 are performed to grasp the needle, we define their reference frame as the needle object frame $\mathcal{F}^{\text{needle}}$ and their goal configuration as ${}^{\text{grasp}}_{\text{needle}}\boldsymbol{T}$. ${}^{\text{needle}}_{\text{endoscope}}\boldsymbol{T}$ is found by segmenting needle pixels by hand, then applying the tracking algorithm of Section 3.3.

Actions A3 and A4 are defined w.r.t. the needle insertion frame $\mathcal{F}^{\text{entry}}$. The transformation ${}^{\text{entry}}_{\text{endoscope}}\boldsymbol{T}$ is found by triangulation of the points observed in the stereo frames.



(a) Initial position (b) Move to grasp needle (A1) (c) Grasp needle (A2) (d) Move to insert (A3) (e) Insert needle (A4)

Figure 5: Human demonstration of needle pickup, positioning and insertion, as part of a surgical suturing task.

Figure 6: Laboratory setup with the da Vinci PSM, stereo camera rig and podium with marked points for needle insertion.

3.2 Evaluation Platform

The needle manipulation task was evaluated on a first-generation da Vinci surgical robot with a dVRK. Only a single PSM was actively used in this work. A 8 mm Large Needle Driver (LND) instrument was used as the PSM end-effector. Fig. 6 shows an overview of the laboratory setup used in the experimental evaluation.

For sensor feedback we used a custom stereo camera rig with two Basler acA2500-20gc cameras each fitted with a Tamron M111FM08 8 mm lens. The baseline was 124 mm and the cameras were positioned approx. 350 mm from the workspace. Camera intrinsics were calibrated using the method of [28]. Endoscopes have a much smaller baseline and field of view than our stereo rig, but are also positioned much closer to the workspace.

The robot-to-camera calibration, ${}^{\text{optical}}_{\text{base}}\boldsymbol{T}$, was determined by manually moving the TCP frame of the manipulator to points on a planar calibration board whose center and orientation served as a common reference frame of the camera and PSM.

We used a small podium, made from a plastic frame with white fabric stretched across it, to provide a different task context from the practice suturing pad used in the demonstration data, as shown in Fig. 5. This setup was chosen to fully control the background and keypoint coloring and thus ease the image segmentation problem.

Communication between system software components was achieved using the Robot Operating System (ROS) middleware.

3.3 Computer Vision-Based Object Tracking

Context is needed to register learned surgical actions to a new environment, i.e., for estimating $\mathscr{F}^{\text{needle}}$ and $\mathscr{F}^{\text{entry}}$. We used computer vision to provide this information since cameras (endoscopes) are usually available during RMIS. Specifically, we used stereo vision for estimating the 6 degrees of freedom (DOF) suturing needle pose and the needle insertion point. The stereo tracking algorithm detailed here is directly transferable to any stereo camera, assuming the intrinsic and extrinsic parameters are known.

To simplify the image segmentation problem we used a green-painted ½-circle suturing needle and marked insertion points in red. We note that this could be a significant limitation in applying the proposed framework in a more clinically realistic setting. However, the vision pipeline is not the main focus of this work and could readily be replaced with one that allows detecting the needle without painting it.

The stereo vision pipeline consists of 1) image rectification; 2) filtering to reduce image noise; 3) color segmentation; and 4) 3D pose estimation. Image pixels are segmented by thresholding in the HSV color space. We estimate the needle pose by first using the method of [29] to fit ellipses to green pixels in each image, with RANSAC to make the method more robust to outliers. Green pixels that are close to the curve of the fitted ellipse are selected as belonging to the needle, providing a dense set of needle pixels. 3D points are then reconstructed by triangulation and the 6 DOF pose of the needle is estimated by fitting a plane to the needle 3D points, yielding a normal vector which we take as the z-axis of the coordinate frame $\mathscr{F}^{\text{needle}}$. The vector between the end points of the needle curve is taken as the y-axis, and the cross product between the two is taken as the negative x-axis of $\mathscr{F}^{\text{needle}}$.

Fig. 7 shows a camera frame with the estimated insertion point locations and needle pose overlaid. Since tracking performance affects the overall performance of the proposed method, we evaluate it in Section 4.1.

3.4 Registering Actions to New Task Contexts

Once the action library has been constructed, and a task has been defined, each step of the task will instantiate one or more actions with task-specific parameters. Fig. 8 shows actions of a needle manipulation task being applied in a new task context. Points of interest are observed and labeled by the tracking system for generalizing actions to it.

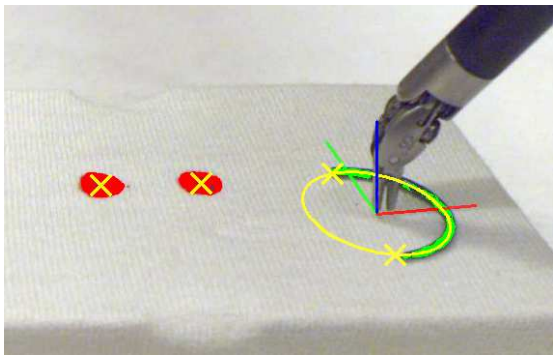


Figure 7: Camera frame with segmentation and needle pose estimation results overlaid.

Initially, the needle is situated in a small block of foam to mimic how it is held by left-hand instrument (PSM2) in the original demonstration. The first action is moving the end-effector into a configuration that lets it grasp the needle. The goal of the action instance is thus a grasp pose, seen w.r.t. $\mathscr{F}^{\text{needle}}$, which is provided by the tracking system. Action A2 is similarly instantiated w.r.t. $\mathscr{F}^{\text{needle}}$. Instantiations of A3 and A4 are both made w.r.t. the needle insertion point in the scene, $\mathscr{F}^{\text{entry}}$.

At a low level, the DMPs of each action are joined by chaining them. This assumes zero velocity at joining points, which is a fair assumption for this task, given that the velocity of demonstration trajectories is close to zero at the beginning and end of the different segmented gestures.

4 Experimental Results

In this section we document experiments undertaken to evaluate the tracking system performance and validate our application of an action library in a needle manipulation task for surgical suturing.

4.1 Tracking System Evaluation

We evaluated the precision of our system by examining the 6 DOF needle pose (reported by the tracking system) w.r.t. the PSM TCP pose (estimated from robot joint angles and kinematic model) while the needle was being held rigidly in the gripper jaws. The needle was autonomously picked up from a location that was unknown to the system beforehand, but with a pre-determined (from demonstration) needle-local grasp pose.

To transform the needle pose reported by the tracking algorithm, ${}^{\text{needle}}_{\text{optical}}\boldsymbol{T}_i$, into the TCP frame at discrete instant i we use ${}^{\text{TCP}}_{\text{base}}\boldsymbol{T}_i$ (computed from robot kinematics) and the camera-to-robot calibration ${}^{\text{optical}}_{\text{base}}\boldsymbol{T}$. Let now $\theta_i \boldsymbol{v}_i$ be the axis-angle representation of the 3×3 rotational part and $\boldsymbol{p}_i$ the 3×1 translation vector of ${}^{\text{needle}}_{\text{TCP}}\boldsymbol{T}_i$. We then define translational and rotational difference measures as

$$d_{\text{translation},i} = \|\boldsymbol{p}_i\|_2$$
$$d_{\text{rotation},i} = |\theta_i|,$$

i.e., the Euclidean norm of $\boldsymbol{p}_i$ and the absolute value of the rotation angle θ_i. In the rotational difference measure we care only about the rotation angle and not the axis of the rotation.

Trial	$d_{\text{translation}}$ [mm]	d_{rotation} [deg]
	std. dev.	std. dev.
1 (N = 234)	0.8	12.2
2 (N = 250)	0.8	13.0
3 (N = 248)	0.7	4.7
Mean	0.8	9.7

Table 2: Relative variance of the needle tracking pose estimate

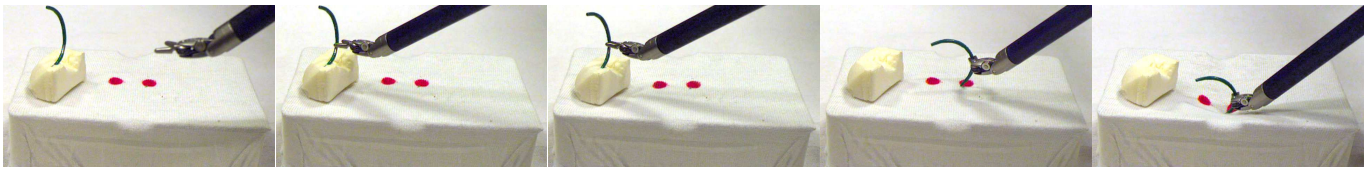
Three trials were performed where the needle was moved to different locations in a 100×100×100 mm volume, which covers the workspace of the autonomous needle manipulation task. On average, 244 corresponding data points were sampled from the tracking system and robot kinematics in each trial. Table 2 shows the standard deviation from the mean of each of the two error measures computed from the collected data points in each trial. Across all samples from all three trials, the standard deviation was 0.8 mm for positional values and 9.7° for rotational values.

4.2 Autonomous Needle Manipulation Task

In this experiment, the system autonomously executed the surgical needle manipulation task, which is detailed in Section 3, in previously unseen task contexts. The needle manipulation actions needed to realize the task were taken from the action library of Section 3.1 and generalized to the new task context, as shown in Fig. 8. The task was repeated in a total of 16 trials. In each trial the needle pickup and insertion goal positions were changed from the demonstration to study how well the surgical actions generalize to new task contexts. Initial positions were chosen by manually moving the fabric-clad podium and needle to cover the visible workspace. The system was initialized with only the action library, the task model and the thresholds for segmenting pixels of specific target objects, which were then automatically labeled to provide the context for generalizing actions.

Fig. 9 plots the demonstration trajectories used to learn actions A1 through A4 as well as the generalized trajectories of A1 through A4 in trials 1, 2, 6 and 7. Action trajectories were scaled temporally by scaling the DMP phase to account for the distance traveled, so that velocities stayed similar to those of the demonstration. Table 3 summarizes the results of the experiment: it shows whether each trial completed successfully[1], how the initial needle position was changed, and the

[1]With "successful completion" we mean that the needle manipulation (grasping and positioning) task was executed without losing the needle.



(a) Initial position (b) Move to grasp needle (A1) (c) Grasp needle (A2) (d) Move to insert (A3) (e) Insert needle (A4)

Figure 8: Autonomous execution of a needle manipulation task based on actions learned from human demonstration (see Fig. 5).

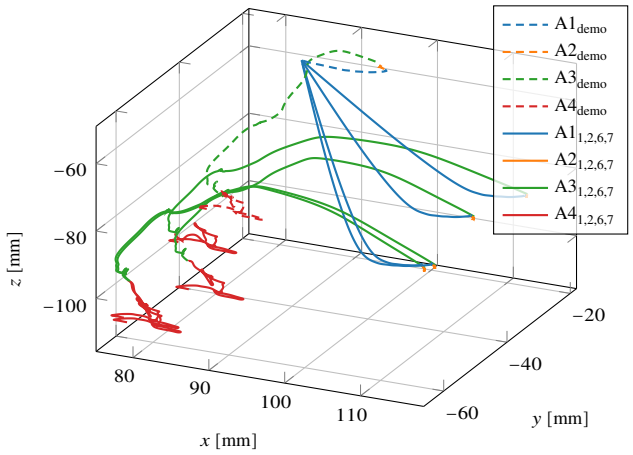


Figure 9: Motions of the original demonstrations and the generalized actions of the needle manipulation task for trials 1, 2, 6 and 7.

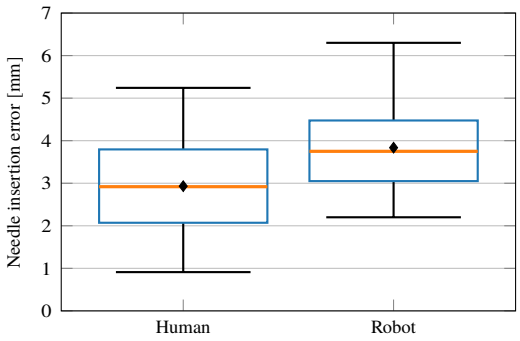


Figure 10: Comparison of human and robot needle insertion errors.

error Euclidean distance between desired and actual insertion points.

To provide a baseline for comparing needle insertion we plot errors of the robot against the those of a human operator in Fig. 10. The human error baseline was established from a set of 13 needle insertion demonstrations, not including the demonstration used for training.

5 Discussion

The experiments in this paper confirm that the presented framework can be used to construct a library of surgical actions. We further show that learned actions can be combined to execute an autonomous needle manipulation task, which generalizes to previously unseen contexts.

Trial	Success[1]	Needle			
		Initial position [mm]			Insert. error [mm]
		x	y	z	
1	Yes	0	0	0	3.7
2	Yes	2.8	−5.3	7.6	3.0
3	Yes	4.2	−16.7	25.3	2.2
4	Yes	3.3	−14.5	21.0	2.7
5	Yes	1.6	−15.5	23.1	6.3
6	Yes	2.6	−16.0	23.9	2.9
7	Yes	1.0	−15.4	24.1	3.2
8	Yes	66.5	12.4	26.2	5.2
9	Yes	55.9	25.1	5.4	4.4
10	Yes	46.5	25.5	4.1	3.3
11	Yes	43.2	26.8	2.4	4.5
12	No	55.9	15.1	24.2	
13	Yes	47.5	9.8	30.9	4.7
14	Yes	25.2	29.4	−2.2	3.8
15	No	21.4	28.6	−1.1	
16	No	8.8	15.3	−29.8	
Mean	81.25 %	24.2	6.5	11.6	3.8

Table 3: Needle manipulation task results. The initial position of the needle is listed relative to its initial position in trial 1.

The current vision-based needle tracking method was robust w.r.t. position, but not w.r.t. orientation, cf. Table 2. Although this error was not caused by the tracking system alone, it did give an upper bound on needle tracking error. The greater variability of errors in the estimated orientations was caused primarily because the needle was partially occluded in a large number of frames.

The needle manipulation task successfully grasped, positioned and pushed down the suturing needle at the designated insertion point in 81 % of 16 trials. In trials 12 and 15 the initial needle grasp failed due to errors in the needle pose estimate. Trial 16 failed because the time scaling constant was set improperly, causing a large acceleration after the initial needle pick-up.

An independent-samples t-test was performed to compare the needle insertion error values of the human user (mean: 2.9 mm, std. dev.: 1.3 mm) and autonomous robot (mean: 3.8 mm, std. dev.: 1.1 mm). There was not nearly enough evidence that the two groups differed significantly, $t(25) = 1.913$, $p = 0.067$, two-tailed, although the human operator performed better than the robot collectively over all trials. For some RMIS procedures an error of 3.8 mm is acceptable although, for some anatomies, it may not be adequate, especially considering the maximum error of 6.3 mm.

A known source of error was the demonstrated needle grasp

pose which was estimated from endoscope images. The endoscope was, however, not calibrated to the PSM, so the calibration had to be estimated. The demonstrations also has the human user grasping the suturing needle about halfway from its tip. Preferably, one would grasp the needle around two thirds toward its end to avoid having to re-grasp it to push it all the way through. Correcting these issues is a matter of using alternative demonstration data that includes camera-to-robot calibration and better grasp pose.

The suturing needle did not fully pierce the surface at the insertion point. This was because the tip of the needle was dulled, due to the paint, and because the fabric surface used in the experiment was very pliant. We used a fabric surface to fully control the color of the background and insertion points. Real or phantom tissue is stiffer and our experiments would likely have been more successful using either of those. Future work is planned to improve vision-based perception, which in turn will allow us to use unmodified needles and more realistic materials.

Initial positions of the suture needle were chosen manually to cover the visible workspace well. They were, however, not chosen systematically. The initial orientation of the needle was kept approximately constant to facilitate pose estimation from vision. Picking up an arbitrarily orientated needle is not currently feasible because of the aforementioned uncertainty of the pose estimation algorithm. The initial position can, in principle, be chosen anywhere in the reachable and visible workspace, but with one caveat: with DMPs, any artifacts of the source demonstration are also present in the execution. Thus, if DMPs are scaled, so are the artifacts.

In the task executions, motions were temporally scaled to maintain velocities similar to those of the original demonstration. Motions could be executed faster, subject to robot limitations, but we have not yet attempted this.

Actual clinical application is not yet realistic both because of ethical and safety concerns and technical limitations. We are confident, though, that we are soon able to robustly automate surgical tasks, like suturing, in ex vivo environments. One major obstacle is perception and another is generating motion. We have planned work to improve vision-based perception algorithms and demonstration data. Furthermore, we plan to extend our framework with the capability to improve already-learned actions from subsequent demonstrations.

6 Conclusion

We have presented a framework for building a surgical action library based entirely on human demonstration. The action library was used to realize a novel application of LfD to a surgical needle manipulation task that was generalized to previously unseen contexts with a success rate of 81 %. The needle manipulation task was executed completely autonomously, using surgical instruments that were not modified in any way. Our approach is independent of the robot system and neither requires a complete description of the environment nor pre-programming motions. Future work is planned to overcome the limitations of our current setup and make possible autonomous execution of more comprehensive surgical tasks based on the proposed framework

Acknowledgment

We thank Giacomo De Rossi, Nicola Piccinelli and Fabio Falezza, all with the University of Verona, for their kind help with the dVRK and in particular robot-to-camera calibration.

References

[1] S. L. Jørgensen, O. Mogensen, C. Wu, K. Lund, M. Iachina, M. Korsholm, and P. T. Jensen, "Nationwide introduction of minimally invasive robotic surgery for early-stage endometrial cancer and its association with severe complications," *JAMA Surgery*, vol. 154, no. 6, p. 530, Jun. 2019.

[2] G.-Z. Yang, J. Cambias, K. Cleary, E. Daimler, J. Drake, P. E. Dupont, N. Hata, P. Kazanzides, S. Martel, R. V. Patel, V. J. Santos, and R. H. Taylor, "Medical robotics—regulatory, ethical, and legal considerations for increasing levels of autonomy," *Science Robotics*, vol. 2, no. 4, Mar. 2017.

[3] P. Kazanzides, Z. Chen, A. Deguet, G. S. Fischer, R. H. Taylor, and S. P. DiMaio, "An open-source research kit for the da vinci® surgical system," in *2014 IEEE Int. Conf. Robot. Autom. (ICRA)*, May 2014, pp. 6434–6439.

[4] N. Padoy and G. D. Hager, "Human-machine collaborative surgery using learned models," in *2011 IEEE Int. Conf. Robot. Autom. (ICRA)*, May 2011, pp. 5285–5292.

[5] J. Schulman, A. Gupta, S. Venkatesan, M. Tayson-Frederick, and P. Abbeel, "A case study of trajectory transfer through non-rigid registration for a simplified suturing scenario," in *2013 IEEE/RSJ Int. Conf. Intell. Robot. Syst. (IROS)*, Nov. 2013, pp. 4111–4117.

[6] C. Staub, T. Osa, A. Knoll, and R. Bauernschmitt, "Automation of tissue piercing using circular needles and vision guidance for computer aided laparoscopic surgery," in *2010 IEEE Int. Conf. Robot. Autom. (ICRA)*, May 2010, pp. 4585–4590.

[7] R. C. Jackson and M. C. Çavuşoğlu, "Needle path planning for autonomous robotic surgical suturing," in *2013 IEEE Int. Conf. Robot. Autom. (ICRA)*, May 2013, pp. 1669–1675.

[8] S. Iyer, T. Looi, and J. Drake, "A single arm, single camera system for automated suturing," in *2013 IEEE Int. Conf. Robot. Autom. (ICRA)*, May 2013, pp. 239–244.

[9] T. Liu and M. C. Çavuşoğlu, "Optimal needle grasp selection for automatic execution of suturing tasks in robotic minimally invasive surgery," in *2015 IEEE Int. Conf. Robot. Autom. (ICRA)*, May 2015, pp. 2894–2900.

[10] S. Sen, A. Garg, D. V. Gealy, S. McKinley, Y. Jen, and K. Goldberg, "Automating multi-throw multilateral surgical suturing with a mechanical needle guide and sequential convex optimization," in *2016 IEEE Int. Conf. Robot. Autom. (ICRA)*, May 2016, pp. 4178–4185.

[11] C. D'Ettorre, G. Dwyer, X. Du, F. Chadebecq, F. Vasconcelos, E. De Momi, and D. Stoyanov, "Automated pick-up of suturing needles for robotic surgical assistance," in *2018 IEEE Int. Conf. Robot. Autom. (ICRA)*, May 2018, pp. 1370–1377.

[12] B. Thananjeyan, A. Tanwani, J. Ji, D. Fer, V. Patel, S. Krishnan, and K. Goldberg, "Optimizing robot-assisted surgery suture plans to avoid joint limits and singularities," in *2019 Int. Symp. Med. Robot. (ISMR)*, Apr. 2019.

[13] S. A. Pedram, C. Shin, P. W. Ferguson, J. Ma, E. P. Dutson, and J. Rosen, "Autonomous suturing framework and quantification using a cable-driven surgical robot," *IEEE Trans. Robot.*, pp. 1–14, 2020.

[14] S. Leonard, K. L. Wu, Y. Kim, A. Krieger, and P. C. W. Kim, "Smart tissue anastomosis robot (star): A vision-guided robotics system for laparoscopic suturing," *IEEE Trans. Biomed. Eng.*, vol. 61, no. 4, pp. 1305–1317, Apr. 2014.

[15] H. Saeidi, H. N. D. Le, J. D. Opfermann, S. Leonard, A. Kim, M. H. Hsieh, J. U. Kang, and A. Krieger, "Autonomous laparoscopic robotic suturing with a novel actuated suturing tool and 3d endoscope," in *2019 Int. Conf. Robot. Autom. (ICRA)*, May 2019.

[16] A. Murali, S. Sen, B. Kehoe, A. Garg, S. McFarland, S. Patil, W. D. Boyd, S. Lim, P. Abbeel, and K. Goldberg, "Learning by observation for surgical subtasks: Multilateral cutting of 3d viscoelastic and 2d orthotropic tissue phantoms," in *2015 IEEE Int. Conf. Robot. Autom. (ICRA)*, May 2015, pp. 1202–1209.

[17] B. Thananjeyan, A. Garg, S. Krishnan, C. Chen, L. Miller, and K. Goldberg, "Multilateral surgical pattern cutting in 2d orthotropic gauze with deep reinforcement learning policies for tensioning," in *2017 IEEE Int. Conf. Robot. Autom. (ICRA)*, May 2017, pp. 2371–2378.

[18] J. van den Berg, S. Miller, D. Duckworth, H. Hu, A. Wan, X.-Y. Fu, K. Goldberg, and P. Abbeel, "Superhuman performance of surgical tasks by robots using iterative learning from human-guided demonstrations," in *2010 IEEE Int. Conf. Robot. Autom. (ICRA)*, May 2010.

[19] A. Knoll, H. Mayer, C. Staub, and R. Bauernschmitt, "Selective automation and skill transfer in medical robotics: A demonstration on surgical knot-tying," *Int. J. Med. Robot. Comput. Assist. Surg.*, vol. 8, no. 4, pp. 384–397, Dec. 2012.

[20] T. Osa, N. Sugita, and M. Mitsuishi, "Online trajectory planning and force control for automation of surgical tasks," *IEEE Trans. Autom. Sci. Eng.*, vol. 15, no. 2, pp. 675–691, Apr. 2018.

[21] P. Berthet-Rayne, M. Power, H. King, and G.-Z. Yang, "Hubot: A three state human-robot collaborative framework for bimanual surgical tasks based on learned models," in *2016 IEEE Int. Conf. Robot. Autom. (ICRA)*, May 2016, pp. 715–722.

[22] B. Kehoe, G. Kahn, J. Mahler, J. Kim, A. Lee, A. Lee, K. Nakagawa, S. Patil, W. D. Boyd, P. Abbeel, and K. Goldberg, "Autonomous multilateral debridement with the raven surgical robot," in *2014 IEEE Int. Conf. Robot. Autom. (ICRA)*, 2014, pp. 1432–1439.

[23] D. Seita, S. Krishnan, R. Fox, S. McKinley, J. Canny, and K. Goldberg, "Fast and reliable autonomous surgical debridement with cable-driven robots using a two-phase calibration procedure," in *2018 IEEE Int. Conf. Robot. Autom. (ICRA)*, May 2018, pp. 6651–6658.

[24] S. Schaal, "Dynamic movement primitives – a framework for motor control in humans and humanoid robotics," in *Adaptive Motion of Animals and Machines*, 2006, pp. 261–280.

[25] A. J. Ijspeert, J. Nakanishi, H. Hoffmann, P. Pastor, and S. Schaal, "Dynamical movement primitives: Learning attractor models for motor behaviors," *Neural Comput.*, vol. 25, no. 2, pp. 328–373, Feb. 2013.

[26] A. Ude, B. Nemec, T. Petric, and J. Morimoto, "Orientation in cartesian space dynamic movement primitives," in *2014 IEEE Int. Conf. Robot. Autom. (ICRA)*, May 2014, pp. 2997–3004.

[27] M. Bombieri, D. Dall'Alba, S. Ramesh, G. Menegozzo, C. Schneider, and P. Fiorini, "Joints-space metrics for automatic robotic surgical gestures classification," in *2020 IEEE/RSJ Int. Conf. Intell. Robot. Syst. (IROS)*, Oct. 2020, pp. 3061–3066.

[28] Z. Zhang, "A flexible new technique for camera calibration," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 11, pp. 1330–1334, 2000.

[29] A. Fitzgibbon, M. Pilu, and R. Fisher, "Direct least square fitting of ellipses," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 21, no. 5, pp. 476–480, May 1999.