Generalization for Deep Reinforcement Learning for Inverse Kinematics of Concentric Tube Robots

K. Iyengar[1], S. Spurgeon[2], and D. Stoyanov[1]

[1]*Department of Computer Science, University College London*
[2]*Department of Electronic and Electrical Engineering, University College London*
keshav.iyengar@ucl.ac.uk

INTRODUCTION

Concentric tube robots (CTRs) are a class of continuum robot that depend on the interactions between neighbouring, concentrically aligned tubes to produce the curvilinear shapes of the robot backbone [1]. The main application of these unique robots is that of minimally invasive surgery (MIS), where most of the developments for CTRs have been focused. Due to the confined workspaces and resulting extended learning times for surgeons in MIS, dexterous, compliant continuum robots such as CTRs have been under development in preference to the mechanically rigid and limited degrees-of-freedom (DOF) robots used in interventional medicine today. The precurved tubes in CTRs, sometimes referred to as active cannulas or catheters, are manufactured from super-elastic materials like Nickel-Titanium alloys with each tube nested concentrically. From the base, the individual tubes can be actuated through extension and rotation, which results in the bending and twisting of the backbone as well as access to the surgical site through the channel and robot tip. Clinically, CTRs are motivated for use in brain, cardiac, gastric surgery as well other procedures [2].

Due to tube interactions, modelling and control is non-trivial. Position control for CTRs has relied on model development, and although a balance between computation and accuracy has been reached in the literature [1], there remain issues such as performance in the presence of tube parameter discrepancies and the impact of unmodelled physical phenomena such as friction and permanent plastic deformation. This motivates the development of an end-to-end model-free control framework for CTRs. We extend our previous model-free deep reinforcement learning (deepRL) method [3] with an initial proof of concept for generalization. The task we give the agent then is to control the end-effector Cartesian robot tip position by means of actions that represent changes in joint values to reach a desired position in the robot workspace whilst considering a specific CTR system.

A hurdle with using deep learning approaches for control of CTRs is the limitation of CTR system generalization. Deep learning methods rely on the data given and cannot

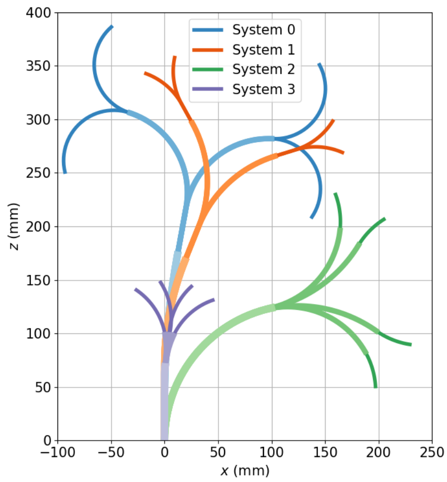


Fig. 1 CTR systems generalized over ordered from longest to shortest.

inherently differentiate between CTR systems. Thus, deep learning methods have been investigated for single CTR systems only. For deep learning methods to be viable, they must be able to generalize over multiple CTR systems.

MATERIALS AND METHODS

The CTR system generic deepRL method described below will seek to generalize over four CTR systems which each have different tube parameters as shown in Fig. 1. These parameters include, stiffness, inner and outer diameters, curved and straight lengths etc. The objective is to obtain good performance across CTR systems with a single control policy. For generalization, a system specifier, $\psi = \{0, 1, 2, 3\}$, was appended to the state, s_t, for the agent to differentiate the CTR systems. During training, a discrete uniform distribution is sampled to determine the system parameters to be used in the simulation for that episode. The simulation, using the exact kinematics from Rucker et al. [4], generates desired goals, G_d, within the workspace of the selected CTR system during the start of an episode and determines the current robot tip position or achieved goal, G_a. During the episode, the agent tries to reach

TABLE I

System	Errors (mm) ± std	Errors (% length) ± std	Success rate
1	0.77 ± 0.65	0.16 ± 0.065	94.0%
2	0.75 ± 0.48	0.17 ± 0.18	94.1%
3	0.63 ± 0.26	0.20 ± 0.08	99.3%
4	0.64 ± 0.22	0.3 ± 0.13	99.3%

this desired goal. Both these goals are included in the state which is defined as

$$s_t = \{\gamma_1, \gamma_2, \gamma_3, G_a - G_d, \delta(t), \psi\} \quad (1)$$

where γ_i is the cylindrical representation [5] for tube i. Tubes are ordered innermost to outermost. The cylindrical representation is defined as:

$$\gamma_i = \{\gamma_{1,i}, \gamma_{2,i}, \gamma_{3,i}\} = \{\cos(\alpha_i), \sin(\alpha_i), \beta_i\} \quad (2)$$

with rotation and extension of tube i represented as α_i and β_i respectively. The agent can take an action a or a change in joint position at each timestep t such that

$$a_t = \{\Delta\beta_1, \Delta\beta_2, \Delta\beta_3, \Delta\alpha_1, \Delta\alpha_2, \Delta\alpha_3\}. \quad (3)$$

The agent receives a reward r_t if the current achieved tip position G_a is within a goal tolerance $\delta(t)$ to the desired tip position G_d. The reward, r_t is defined as:

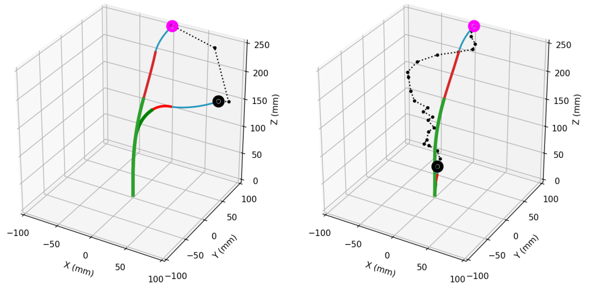
$$r_t = \begin{cases} 0 & \text{if } e_t \leq \delta(t) \\ -1 & \text{otherwise} \end{cases}. \quad (4)$$

where e_t is the Euclidean distance or l_2 norm $||G_a - G_d||_2$ between the achieved and desired goal.

Using deep deterministic policy gradient (DDPG) [6] with hindsight experience replay (HER) [7], the training parameters are as follows. The number of training timesteps was 3 million, buffer size was $500,000$ with the policy network having 3 hidden networks with 256 units per layer, the initial goal tolerance and final goal tolerance were 20 mm and 1 mm applied over 1.5 million steps using a decay function [3]. Zero-mean Gaussian noise of 1.8 mm was applied to $\Delta\beta_i$ and 0.025 radians to $\Delta\alpha_i$.

RESULTS

A generic policy for all four systems was trained then using the trained policy, 1000 evaluation episodes were performed for each CTR system, resulting in 4000 evaluation episodes in total. For training results, a success rate of 100% was achieved with a mean error of 0.7 mm in the final 100 episodes. For each respective system, the mean and standard deviation errors were 0.77 mm and 0.65 mm, 0.75 mm and 0.48 mm, 0.63 mm and 0.26 mm and 0.64 mm and 0.22 mm. The success rate for each system was 94.0%, 94.1% 99.3% and 99.3%. These results are summarized in Table I. As longer CTR systems will have larger errors, usually error metrics are represented as a percentage of robot length. Even still, we find larger errors found in robot systems with larger workspaces.



(a) (b)

Fig. 2 System 0 (a) and system 1 (b) inverse kinematic solutions for the same desired goal.

Visualized in Fig. 2 is an example of the same desired end-effector position with two different initial joint configurations, resulting in two different final inverse kinematics solutions for system 0 Fig. 2a and system 1 Fig. 2b using the same policy. The desired goal position was $(0, 100, 250)$ mm. The final joint configuration in Figure 2a was $\beta = [-7.63, -4.78, 0.0]$ mm, $\alpha = [65.83, 200.19, 120.77]°$ with a tip error of 0.58 mm and for Fig. 2b $\beta = [-4.72, -3.54, -0.09]$ mm and $\alpha = [62.61, 193.39, 172.42]°$ with a tip error of 0.65 mm.

DISCUSSION

The method is able to generalize over four distinct CTR systems however evaluation metrics across systems differ as seen in Table I. In other words, error metrics is biased to smaller robot workspaces even when represented as a percentage of robot length. System 0, the system with the longest overall length performs worst in evaluation whereas system 3 performs best. In the future, we plan on including CTR system parameters to fully generalize.

REFERENCES

[1] H. B. Gilbert, D. C. Rucker, and R. J. Webster III, "Concentric tube robots: The state of the art and future directions," *Robotics Research*, pp. 253–269, 2016.

[2] H. Alfalahi, F. Renda, and C. Stefanini, "Concentric tube robots for minimally invasive surgery: Current applications and future opportunities," *IEEE Transactions on Medical Robotics and Bionics*, vol. 2, no. 3, pp. 410–424, 2020.

[3] K. Iyengar and D. Stoyanov, "Deep reinforcement learning for concentric tube robot control with a goal-based curriculum," in *2021 IEEE International Conference on Robotics and Automation (ICRA)*, 2021, pp. 1459–1465.

[4] D. C. Rucker, B. A. Jones, and R. J. Webster III, "A geometrically exact model for externally loaded concentric-tube continuum robots," *IEEE transactions on robotics*, vol. 26, no. 5, pp. 769–780, 2010.

[5] R. Grassmann, V. Modes, and J. Burgner-Kahrs, "Learning the forward and inverse kinematics of a 6-dof concentric tube continuum robot in se(3)," in *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2018, pp. 5125–5132.

[6] T. P. Lillicrap, J. J. Hunt, A. Pritzel, N. Heess, T. Erez, Y. Tassa, D. Silver, and D. Wierstra, "Continuous control with deep reinforcement learning," *arXiv preprint arXiv:1509.02971*, 2015.

[7] M. Andrychowicz, F. Wolski, A. Ray, J. Schneider, R. Fong, P. Welinder, B. McGrew, J. Tobin, P. Abbeel, and W. Zaremba, "Hindsight experience replay," *arXiv preprint arXiv:1707.01495*, 2017.