2020 IEEE International Conference on Robotics and Automation (ICRA)
31 May - 31 August, 2020. Paris, France

Reinforcement Learning Based Manipulation Skill Transferring for Robot-assisted Minimally Invasive Surgery

Hang Su[1], Yingbai Hu[2*], Zhijun Li[3], Alois Knoll[2], Giancarlo Ferrigno[1] and Elena De Momi[1]

Abstract— **The complexity of surgical operation can be released significantly if surgical robots can learn the manipulation skills by imitation from complex tasks demonstrations such as puncture, suturing, and knotting, etc.. This paper proposes a reinforcement learning algorithm based manipulation skill transferring technique for robot-assisted Minimally Invasive Surgery by Teaching by Demonstration. It employed Gaussian mixture model and Gaussian mixture Regression based dynamic movement primitive to model the high-dimensional human-like manipulation skill after multiple demonstrations. Furthermore, this approach fascinates the learning and trial phase performed offline, which reduces the risks and cost for the practical surgical operation. Finally, it is demonstrated by transferring manipulation skills for reaching and puncture using a KUKA LWR4+ robot in a lab setup environment. The results show the effectiveness of the proposed approach for modelling and learning of human manipulation skill.**

I. INTRODUCTION

Robotic surgery has been a compelling emerging technology that holds significant promise due to the benefits it provides for surgeons, such as higher operational accuracy, extended motion range, and augmented visualization [1], [2]. However, due to the kinematic constraints imposed by the laparoscopic surgery, i.e., which are known as the remote center of motion (RCM) constraints, the surgical operation is performed in a limited space [3]. It turns the intuitive manipulation in the conventional open surgery to time-consuming tasks [4]. The complexity of surgical operation can be released significantly if surgical robots can learn the manipulation skills by imitation from complex tasks demonstrations [5] such as puncture and knotting, etc..

With the development of technology in artificial intelligence and cognition progress, increasing the autonomy of surgical robots in performing some specific complex surgical operations, like suturing or knotting, can potentially reduce the length of surgical procedures and surgeon fatigue, as well as improved accuracy [6]. Hence, the need for developing methodology and technology in surgical manipulation skill transferring reinforces.

Teaching by demonstration (TbD) has drawn extensive research attention in manipulation skill transferring from human to robot during the past decades [7], [8]. Calinon *et al.* [9], [10] investigated the methods to assign human motion skills to the robot manipulators. Dynamic movement primitive (DMP) proposed by Meier *et al.* [11] is an efficient approach to learn motor primitives for robot manipulation. In the motion modeling paradigm, each manipulation procedure features motion primitives and the corresponding goal parameters. Kormushev *et al.* [12] studied comprehend the trajectory generation for the spherical impediment by using DMP modelling combined with the synthetic capacity discipline method using Gaussian mixture Regression (GMR). For complex manipulations modelling, the reinforcement learning algorithm is capable of adapting the goal parameters with a high-dimension motion primitive [13], [14].

It is interesting to introduce GMR based motion primitives modeling strategy to learn the surgical operation skills from experienced surgeons [15]. Furthermore, it is proposed to handle the RCM constraint as sub-goal by offline tasks learning and trials using reinforcement learning. In this paper, a special reinforcement learning named Policy Improvement with Path Integrals ($\mathbf{PI^2}$) is proposed to optimize the path planing, which is derived from probability-based stochastic optimal control theory [13], [16]. The reinforcement learning method can optimize the trajectory with disturbance by updating the parameters [17]. In addition, we can design the cost function to explore the different learning tasks even multi-tasks at a learning system. The $\mathbf{PI^2}$ is suitable for the high dimensions problem such as Cartesian space [18]. Therefore, it is convenient for robot learning problem.

It must be noticed that the proposed methodology represents an improvement with respect to the simple path planning between the start and end-point introduced in [13], [19], and combines an sub-goal task to respect the RCM constraint [20], [21], [2] on the planned path. It means the planned path should consider not only the shape but also passing through the small incisions on the abdominal wall. Furthermore, it fascinates the learning and trial phase performed offline, which reduces the risks and cost for the practical surgical operation. Finally, experiments have been performed to demonstrate the proposed control method on a 3-D printed patient phantom using a 7-DoF robot manipulator KUKA LWR4+ .

*This work was supported in part by the European Commission Horizon 2020 research and innovation program, under the project SMARTsurg, grant agreement No. 732515 and in part by the Human Brain Project SGA2, under the Specific Grant Agreement No. 785907.

*Corresponding Author

[1]Hang Su, Giancarlo Ferrigno and Elena De Momi are with the Dipartimento di Elettronica, Informazione e Bioingegneria, Politecnico di Milano, 20133, Milano, Italy. {hang.su, giancarlo.ferrigno, elena.demomi}@polimi.it

[2]Yingbai Hu and Alois Knoll are with the Department of Informatics, Technical University of Munich, Munich, 85748, Germany. yingbai.hu@tum.de; knoll@mytum.de

[3]Zhijun Li is with Department of Automation, University of Science and Technology of China, China. zjli@ieee.org

II. Motivation and Previous Works

In our previous works [2], [20], [21], we inserted the surgical tool into the abdominal cavity passing through the RCM constraint, by hands-on control. To reduce the complexity of operation procedures, it is interesting to utilize the TbD techniques to model and learning the surgeons' manipulation skill and transfer it to the robot, increasing the autonomy of surgical robots.

In our previous work, we utilized the reinforcement learning to learn the complex motion sequences in human-robot environment such that the robots can adapt its motions for manipulation and grasping of a mobile manipulator [19]. Nevertheless, the above algorithms considered only point to point problems, which determine the trajectory between the start and end-point of the movement, ignoring the other goals in the sub-tasks [13], such as passing through a kinematic constraint. In this paper, it is suggested to include the RCM constraint as a new challenge for the manipulation skill modelling and transferring. We therefore extend the policy improvement with path integrals ($\mathbf{PI^2}$) algorithm to simultaneously optimize shape and goal parameters.

III. Control Methodology

The control methodology here proposed aims to provide consistent and effective skill modelling and transferring techniques for robot-assisted minimally invasive surgery, learning the motion primitives of a specific task from demonstration operations operated by a surgeon, and plan the path with respect to the kinematic constraint (RCM constraint) between the start and end-point. The robot model has been discussed in [20], [2].

A. Dynamic Movement Primitive

Given the continuous stream of movements that biological systems exhibit in their daily activities, dynamic movement primitive [19] now is a general approach in artificial and biological systems revolves around identifying movement primitives for motor control in robotics and biology. Dynamic movement primitive is represented as a set of equations, and it can model different linear or nonlinear motions which is convenient to imitate learning complex movement fusion with reinforcement learning algorithm. The dynamic movement primitive is expressed as:

$$\begin{aligned} &\ddot{X}_t = K_p\left(g - X_t\right) - K_v \dot{X}_t + F\left(s_t\right) \\ &\dot{s}_t = \alpha_s s_t \\ &F\left(s_t\right) = h_t^{\mathrm{T}}\left(s_t\right) \omega \left(g - X_0\right) \end{aligned} \tag{1}$$

$$h_t\left(s_t\right) = \frac{\sum\limits_{i=1}^{N} \psi_i\left(s_t\right) s_t}{\sum\limits_{i=1}^{N} \psi_i(s_t)_t}, \psi_i\left(s_t\right) = \exp\left(-\frac{1}{2\sigma_i}\left(s_t - c_i\right)^2\right)$$

where $\left[X_t, \dot{X}_t, \ddot{X}_t\right]$ is the Cartesian space trajectory; X_0 and g present the initial position and goal position of the attractor point in Cartesian space, respectively; K_p and K_d are the stiffness matrix, damping term of DMP in 3D Cartesian space. ω is the shape parameter of DMP; α_s is the scale parameter of Canonical system, where s_t asymptotically decays from 1 to 0; σ_i and c_i are bandwidth and center of the i-th Gaussian kernels.

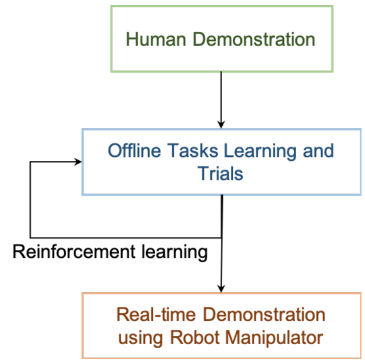

Fig. 1. Experimental procedures.

It should be noted that DMP is consist of two parts including linear spring damper part $K_p\left(g - X_t\right) - K_v \dot{X}_t$ and nonlinear part $F\left(s_t\right)$ which can be applied to model the trajectories from teaching by demonstrations even the nonlinear system. Therefore, DMP is convenient to imitate the motions from human because of the feature of convergence to the attract point g.

B. Gaussian Mixture Model

In this part, the Gaussian Mixture Model is presented to encode the trajectories from teaching. Gaussian Mixture Model is a probability-based statistical model which can describe the probability density distribution of high-dimensional dataset by the sum of different weights of multiple Gaussian models [22]. In this paper, the GMM is used to describe the position density in Cartesian space and obtain nonlinear item in in DMP by regression from each GMM. The DMP framework of multi-demonstrations is reformulated as by K component Gaussian model,

$$\ddot{X} = \sum_{k=1}^{K} h_k \left(K_k^p \left(\mu_k^X - X\right) - K_k^V \dot{X} + F\right) \tag{2}$$

The Cartesian space data point from demonstrations are defined as: $s_j = \left(s_{t,j}, s_{X,j}\right)(j = 1, \ldots, N)$, where N is the length of dataset. Each datapoint include the time temporal value $s_{t,j}$ and position value $s_{X,j}$. To encode the dataset of position distribution $P(s_t, s_X)$, the following GMM model is defined as

$$p\left(s_j\right) = \sum_{k=1}^{K} p\left(k\right) p\left(s_j | k\right) \tag{3}$$

where K is the number of the Gaussian model; $p\left(k\right)$ denotes the prior probability, and $p\left(s_j | k\right)$ is the conditional probability density function.

The manipulator works in 3-D space, so the parameters in (3) are denoted as

$$\begin{aligned} & p(k)=\lambda_k \\ & p(s_j \mid k)=\frac{1}{\sqrt{(2\pi)^3\left|\sum_k\right|}} \mathrm{e}^{\left(-\frac{1}{2}(s_j-\mu_k)^{\mathrm{T}} \sum_k^{-1}(s_j-\mu_k)\right)} \end{aligned} \tag{4}$$

We define the GMM parameters as $\Theta=\{\lambda_k, \mu_k, \sum_k, E_k\}$, where λ_k, μ_k, $\sum_k$, E_k are prior probability, mean variable, covariance variable and cumulated posterior probability, respectively. According to Bayes theorem, the cumulated posterior probability E_k can be expressed as,

$$\begin{aligned} & E_k=\sum_{j=1}^N p(k \mid s_j) \\ & p(k \mid s_j)=\frac{p(k) p(s_j \mid k)}{\sum\limits_{m=1}^K p(m) p(s_j \mid m)} \end{aligned} \tag{5}$$

Then, the log-likelihood of GMM model Θ is defined,

$$\mathcal{L}_{\Theta}=\frac{1}{N} \sum_{j=1}^N \log(p(s_j)) \tag{6}$$

where $p(s_j)=\sum\limits_{k=1}^K p(k) P(s_j \mid k)$. To estimate GMM parameters $\Theta=\{\lambda_k, \mu_k, \sum_k, E_k\}$, the EM algorithm is proposed to train the model parameters, and we will obtain the model parameters after the parameters convergence. The iteration finished step is set when $\frac{\mathcal{L}_{\Theta}^{(t+1)}}{\mathcal{L}_{\Theta}^{(t)}} \leq 0.01$.

C. Gaussian Mixture Regression

Actually, the aim of training is to get the regression parameter F from the dataset. After the multi-demonstrations probability distributions is obtained by GMM, then the Gaussian Mixture Regression (GMM) is proposed to reconstruct the general form for the dataset.

To estimate the conditional expectation value, the observations parameters is defined as: $s=\{s_t, s_X\}$ where s_X is the spatial variable at time step s_t. So, the goal of regression is to estimate the conditional expectation of s_X when the time step s_t is fist given.

For multi-demonstrations from teaching, the GMM Θ encodes the set of trajectories from robot in Cartesian space. The k component of Gaussian mixture model is defined as,

$$\mu_k=\{\mu_{t,k}, \mu_{X,k}\} \quad , \quad \Sigma_k=\begin{pmatrix} \Sigma_{tt,k} & \Sigma_{tX,k} \\ \Sigma_{Xt,k} & \Sigma_{XX,k} \end{pmatrix} \tag{7}$$

where μ_k and Σ_k are mean and covariance matrix of k component GMM. When the time step s_t is given, the expected distribution $s_{X,k}$ of k-th component is expressed as,

$$\begin{aligned} p(s_{X,k} \mid s_t, k) &= \mathcal{N}\left(s_{X,k} ; \hat{s}_{X,k}, \hat{\Sigma}_{XX,k}\right) \\ \hat{s}_{X,k} &= \mu_{X,k}+\sum_{Xt,k}(\Sigma_{tt,k})^{-1}(s_t-\mu_{t,k}) \\ \hat{\Sigma}_{XX,k} &= \Sigma_{XX,k}-\Sigma_{Xt,k}(\Sigma_{tt,k})^{-1} \Sigma_{tX,k} \end{aligned} \tag{8}$$

where $\hat{s}_{X,k}$ and $\hat{\Sigma}_{XX,k}$ are mixed from probability. According to the GMM parameters $\Theta=\{\lambda_k, \mu_k, \sum_k, E_k\}$, the condition probability density is obtained as,

$$p(s_X \mid s_t)=\sum_{k=1}^K h_k \mathcal{N}\left(s_X ; \hat{s}_{X,k}, \hat{\Sigma}_{XX,k}\right) \tag{9}$$

$$h_k=\frac{p(k) p(s_t \mid k)}{\sum_{i=1}^K p(i) p(s_t \mid i)}=\frac{\lambda_k \mathcal{N}(s_t ; \mu_{t,k}, \Sigma_{tt,k})}{\sum_{i=1}^K \lambda_i \mathcal{N}(s_t ; \mu_{t,i}, \Sigma_{tt,i})}$$

From (8) and (9), the estimation of condition expectation s_X and covariance matrix are concluded as,

$$\hat{s}_X=\sum_{k=1}^K h_k \hat{s}_{X,k} \quad , \quad \hat{\Sigma}_{XX}=\sum_{k=1}^K h_k^2 \hat{\Sigma}_{XX,k} \tag{10}$$

Therefore, the motion $\hat{s}=\{\hat{s}_t, \hat{s}_X\}$ can be generated by estimating $\left\{\hat{s}_X, \hat{\Sigma}_{XX}\right\}$ at time step s_t.

D. Reinforcement Learning for Trajectories optimization

In this section, the reinforcement learning is proposed to learning trajectories by learning the shape parameter ω with the noise added. The cost function of Reinforcement learning is defined as,

$$S(T_i)=\phi_{t_N}+\int_{t_i}^{t_N}(r_t+\frac{1}{2} \omega^T R \omega) dt \tag{11}$$

where T_i denotes the trajectory. The cost function S is consist of three parts: terminal cost ϕ_{t_N}, immediate cost r_t, and immediate control $\frac{1}{2} \omega^T R \omega$.

If the noise is added to the shape parameter of DMP $(\omega+\epsilon_t)$, the trajectories would deviation from expectations trajectories. Therefore, the reinforcement learning is applied to learn the shape ω parameter from random noise, and the cost function is reformulated as,

$$S(T_i)=\phi_{t_N}+\int_{t_i}^{t_N} r_t+\frac{1}{2}(\omega+M_t \epsilon_t)^{\mathrm{T}} R(\omega+M_t \epsilon_t) \tag{12}$$

where M_t indicates projection matrix onto the range space of h_t which is defined as,

$$M_t=\frac{R^{-1} h_t h_t^T}{h_t^{\mathrm{T}} R^{-1} h_t}$$

The learning system is aim at minimizing the cost function S by learning the shape parameter ω.

According to the stochastic optimal control theory [13], the path integral of cost function is defined as,

$$\delta \omega_t=\int P(T_i) M_t d T_i \tag{13}$$

where $p(T_i)$ is the probability of trajectory T_i and it is expressed as,

$$P(T_i)=\frac{\exp\left(-\frac{1}{\gamma} S(T_i)\right)}{\int \exp\left(-\frac{1}{\gamma} S(\tau_i)\right) d T_i}$$

Then, the change of $\delta \omega$ can be concluded as,

$$[\delta \omega]_j=\frac{\sum_{i=0}^{N-1}(N-i) w_{j,t_i}[\delta \omega_{t_i}]_j}{\sum_{i=0}^{N-1} w_{j,t_i}(N-i)} \tag{14}$$

where $[\delta\omega]_j$ denotes the j-th element of shape parameter ω. Finally, the new parameters can be obtained,

$$\omega^{new} = \omega + \delta\omega \tag{15}$$

The update rule of reinforcement learning is shown in $\mathbf{Algorithm.1}$.

Algorithm 1 :Reinforcement Learning Update Rule

Initialization:
r_t, ϕ_{t_N}, $f(s_t)$, ω_0, initial state X_0.
While stop after cost convergence
1. **For** $n = 1, \ldots, N_k$,
   - sample from dataset with random noise.
     $\epsilon_{t,n} \sim \mathcal{N}(0, \sigma^2 \mathbf{\Sigma})$
   - compute the cost and probability.
     $S_{T_i,n} = S(\omega + \epsilon_{t,n})$
     $P_{T_i,n} = \frac{\exp\left(-\frac{1}{\gamma} S_{T_i,n}\right)}{\int \exp\left(-\frac{1}{\gamma} S_{T_i,n}\right) dT_i}$
2. **Update mean:**
   $\delta\omega_t = \sum_{n=1}^{N_k} P_{T_i,n} \epsilon_{t,n}$.
   $[\delta\omega]_j = \frac{\sum_{i=0}^{N-1} (N-i) w_{j,t_i} \left[\delta\omega_{t_i}\right]_j}{\sum_{i=0}^{N-1} w_{j,t_i} (N-i)}$
3. **Update parameters:** $\omega^{new} = \omega + \delta\omega$.

IV. Experimental Demonstration

The procedure of the demonstration divided into three phases, shown in Fig. 1 , including demonstration phase, reinforcement learning based offline tasks learning and trials, and reinforcement learning based real-time demonstration.

A. Demonstration Phase

A 3-D printed patient phantom served as the abdomen cavity are used for demonstration, shown in Fig. 2. The surgeon uses hands-on control to insert the surgical tool into the surgical cavity with a repetition of 7 times from different initial point.

Then, dynamic movement primitive is applied to encode the multi-demonstrations from human teaching. The regression results are shown in Figs. 3-5. Fig. 3 shows that all the reproductions can pass through the original RCM and converge to the goal point even from random initial position.

B. Offline Tasks Learning and Trials

Considering the disturbances (goal g and shape ω noise) in the environment, reinforcement learning is applied to optimize the trajectory. The immediate cost is designed as,

$$r_t = \eta_1 \|d_X\| + \eta_2 \left\|\ddot{X}_t\right\|$$

where d_X is the distance from the end-effector to the line connecting start and end-point of the movement. The learning results are shown in Fig. 6. As the number of iterations increases, all the samples gradually converge to the original trajectory with random noise which proves the effectiveness of the proposed methods. Furthermore, RCM constraint is usually not fixed according the actual surgical scenario. Therefore, we hope the manipulator also can learn how to pass through the new RCM for the same tasks without human demonstrations. Hence, the demand to respect the new RCM constraint is treated as sub-goal. Reinforcement learning based offline training and trials are implemented to enable the manipulator to pass through the new rcm constraint without clinical trial. The significance of offline training and trials is to reduce machine wear and avoid dangers in actual experiments.

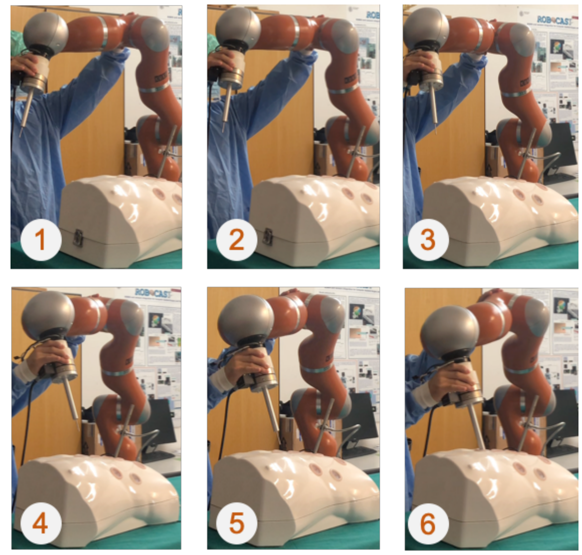

Fig. 2. Demonstration of puncture procedure. The numbers (1-6) indicate the puncture procedure by hands-on demonstration. The 1st picture shows the start point of the tracking tasks, and the 6th picture represents the corresponding end point. The "surgeon" use hands to hold on end-effector and insert the tool tip into the abdominal cavity.

The learning and trials results with new RCM constraint are shown in Figs. 7-8, which prove that the reinforcement learning enhanced $\mathbf{PI^2}$ can adapt the motion skill to meet the new task requirement without re-teaching.

It should be mentioned that in the new RCM task, the immediate cost is re-designed as,

$$r_t = \eta_1 \|d_X\| + \eta_2 \left\|\ddot{X}_t\right\| + \eta_3 \|X_t - P_{RCM}^{new}\|$$

P_{RCM} denotes the manipulator passing through RCM which is set by user. The parameters $K_p = diag[1,1,1,1,1,1,1]$, $K_v = \sqrt{2K_p}$; $\alpha_s = 0.01$; the components of GMM is set $K = 10$; $\eta_1 = 10^6$, $\eta_2 = 10^3$, $\eta_3 = 10^{10}$. $P_{RCM} = [-0.438; 0.4349; 0.2429]m$, the new $P_{RCM}^{new} = [-0.345; 0.4342; 0.2521]m$.

C. Real-time Demonstration with Robot Manipulator

After the offline learning process, the learning results will be applied to demonstrate in actual experiment. The robot perform the insertion of the surgical tool into the abdominal cavity autonomously. Fig. 10 shows one of the demonstrated experiment using KUKA LWR4+ robot manipulator and Fig. 11 shows its performed trajectory.

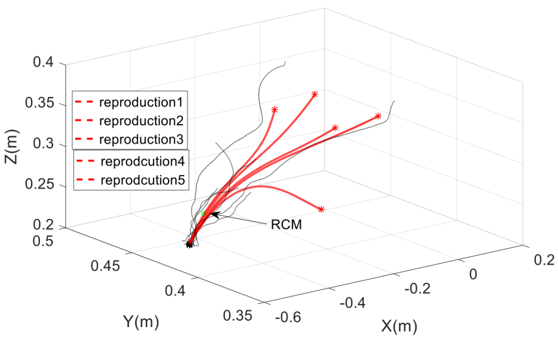

Fig. 3. Reproduction from random initial position via RCM tasks by multi-demonstrations using GMM-GMR. The black curves denote the multi-demonstrations.

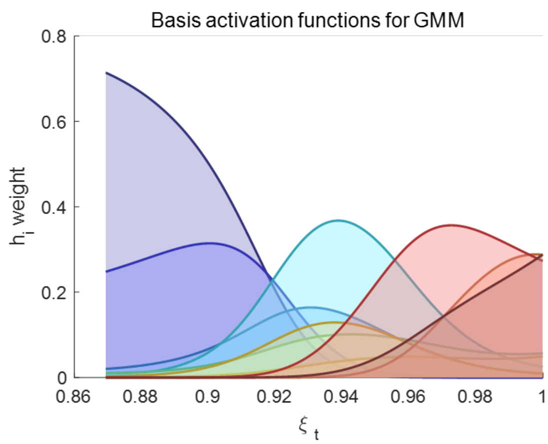

Fig. 4. Basic function of GMM.

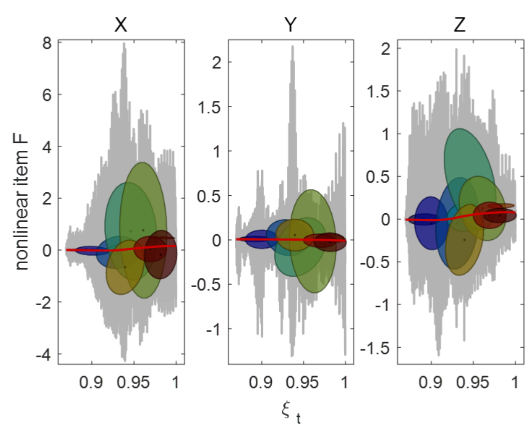

Fig. 5. Regression result of nonlinear term of motion primitives using GMM-GMR from multi-demonstrations.

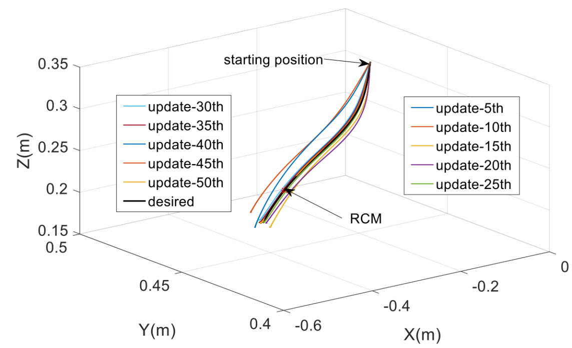

Fig. 6. The learning process of trajectory via original RCM point.

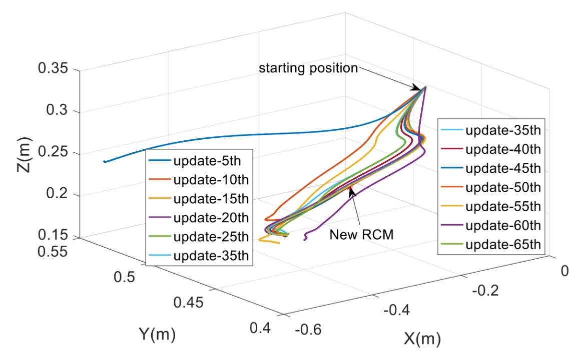

Fig. 7. The learning process of trajectory via new RCM point.

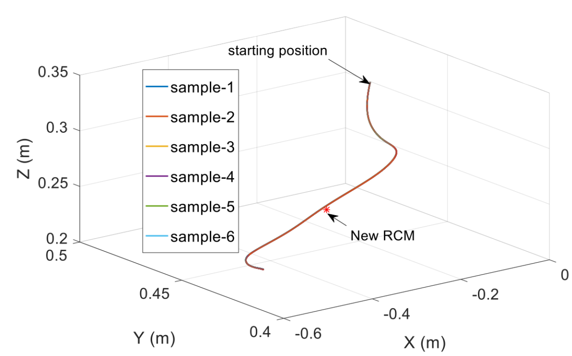

Fig. 8. The last learning results of all samples via new RCM point.

V. Conclusion

This paper proposes a reinforcement learning algorithm based manipulation skill transferring technique for robot-assisted Minimally Invasive Surgery by Teaching by Demonstration. This approach fascinates the learning and trial phase performed offline, which reduces the risks and cost for the practical surgical operation. Reinforce learning is adopted to model the manipulation skill with trials offline until it can handle the varying kinematic constraints. The results have demonstrated the effectiveness of the proposed approach for modelling and learning of human manipulation skill. However, this work considers only kinematic constraints, ignoring the force from physical interaction on the abdominal wall [2]. Future works will involve physical interaction analysis.

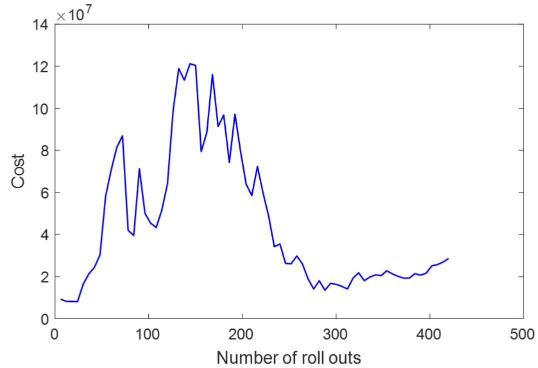

Fig. 9. Cost function by iteration of new RCM tasks.

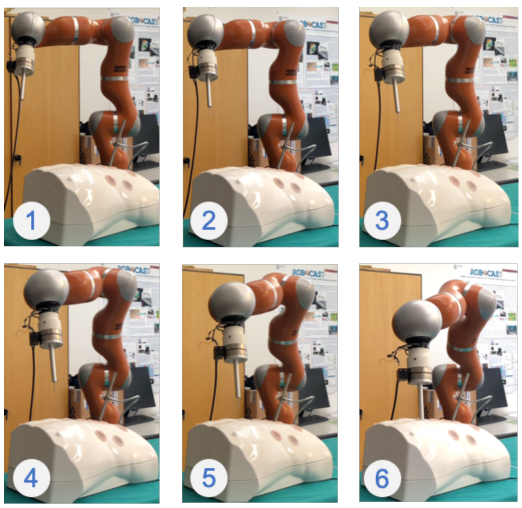

Fig. 10. Demonstration of autonomous puncture using reinforcement learning. The numbers (1-6) indicate the puncture procedure.

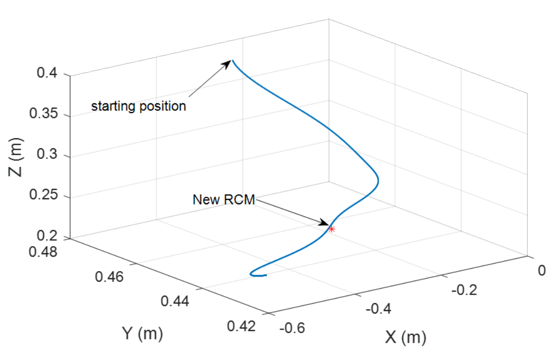

Fig. 11. Demonstrated trajectory curve.

References

[1] A. M. Okamura, C. Simone, and M. D. O'leary, "Force modeling for needle insertion into soft tissue," *IEEE transactions on biomedical engineering*, vol. 51, no. 10, pp. 1707–1716, 2004.

[2] H. Su, C. Yang, G. Ferrigno, and E. De Momi, "Improved human-robot collaborative control of redundant robot for teleoperated minimally invasive surgery," *IEEE Robotics and Automation Letters*, vol. 4, no. 2, pp. 1447–1453, 2019.

[3] H. Su, W. Qi, C. Yang, J. Sandoval, G. Ferrigno, and E. De Momi, "Deep neural network approach in robot tool dynamics identification for bilateral teleoperation," *IEEE Robotics and Automation Letters*, pp. 1–1, 2020.

[4] R. H. Taylor, A. Menciassi, G. Fichtinger, P. Fiorini, and P. Dario, "Medical robotics and computer-integrated surgery," in *Springer handbook of robotics*. Springer, 2016, pp. 1657–1684.

[5] G. A. Fontanelli, M. Selvaggio, L. R. Buonocore, F. Ficuciello, L. Villani, and B. Siciliano, "A new laparoscopic tool with in-hand rolling capabilities for needle reorientation," *IEEE Robotics and Automation Letters*, vol. 3, no. 3, pp. 2354–2361, 2018.

[6] D. Hu, Y. Gong, B. Hannaford, and E. J. Seibel, "Semi-autonomous simulated brain tumor ablation with ravenii surgical robot using behavior tree," in *2015 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2015, pp. 3868–3875.

[7] M. Capurso, M. M. G. Ardakani, R. Johansson, A. Robertsson, and P. Rocco, "Sensorless kinesthetic teaching of robotic manipulators assisted by observer-based force control," in *2017 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2017, pp. 945–950.

[8] Y. Huang, L. Rozo, J. Silvério, and D. G. Caldwell, "Non-parametric imitation learning of robot motor skills," in *2019 International Conference on Robotics and Automation (ICRA)*. IEEE, 2019, pp. 5266–5272.

[9] S. Calinon, D. Bruno, M. S. Malekzadeh, T. Nanayakkara, and D. G. Caldwell, "Human–robot skills transfer interfaces for a flexible surgical robot," *Computer methods and programs in biomedicine*, vol. 116, no. 2, pp. 81–96, 2014.

[10] J. Silvério, S. Calinon, L. Rozo, and D. G. Caldwell, "Learning task priorities from demonstrations," *IEEE Transactions on Robotics*, vol. 35, no. 1, pp. 78–94, 2018.

[11] F. Meier, E. Theodorou, F. Stulp, and S. Schaal, "Movement segmentation using a primitive library," in *2011 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE, 2011, pp. 3407–3412.

[12] P. Kormushev, S. Calinon, and D. G. Caldwell, "Imitation learning of positional and force skills demonstrated via kinesthetic teaching and haptic input," *Advanced Robotics*, vol. 25, no. 5, pp. 581–603, 2011.

[13] F. Stulp, E. A. Theodorou, and S. Schaal, "Reinforcement learning with sequences of motion primitives for robust manipulation," *IEEE Transactions on robotics*, vol. 28, no. 6, pp. 1360–1370, 2012.

[14] O. Koç, G. Maeda, and J. Peters, "Optimizing the execution of dynamic robot movements with learning control," *IEEE Transactions on Robotics*, 2019.

[15] Y. Huang, L. Rozo, J. Silvério, and D. G. Caldwell, "Kernelized movement primitives," *The International Journal of Robotics Research*, vol. 38, no. 7, pp. 833–852, 2019.

[16] P. Englert and M. Toussaint, "Combined optimization and reinforcement learning for manipulation skills." in *Robotics: Science and systems*, vol. 2016, 2016.

[17] S. Parisi, S. Ramstedt, and J. Peters, "Goal-driven dimensionality reduction for reinforcement learning," in *2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2017, pp. 4634–4639.

[18] J. Kober, A. Wilhelm, E. Oztop, and J. Peters, "Reinforcement learning to adjust parametrized motor primitives to new situations," *Autonomous Robots*, vol. 33, no. 4, pp. 361–379, 2012.

[19] Z. Li, T. Zhao, F. Chen, Y. Hu, C.-Y. Su, and T. Fukuda, "Reinforcement learning of manipulation and grasping using dynamical movement primitives for a humanoidlike mobile manipulator," *IEEE/ASME Transactions on Mechatronics*, vol. 23, no. 1, pp. 121–131, 2017.

[20] H. Su, J. Sandoval, M. Makhdoomi, G. Ferrigno, and E. De Momi, "Safety-enhanced human-robot interaction control of redundant robot for teleoperated minimally invasive surgery," in *2018 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2018, pp. 6611–6616.

[21] H. Su, S. Li, J. Manivannan, L. Bascetta, G. Ferrigno, and E. De Momi, "Manipulability optimization control of a serial redundant robot for robot-assisted minimally invasive surgery," in *2019 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2019.

[22] L. Rozo, S. Calinon, D. G. Caldwell, P. Jimenez, and C. Torras, "Learning physical collaborative robot behaviors from human demonstrations," *IEEE Transactions on Robotics*, vol. 32, no. 3, pp. 513–527, 2016.