

# Real-Time Constrained 6D Object-Pose Tracking of An In-Hand Suture Needle for Minimally Invasive Robotic Surgery

Zih-Yun Chiu<sup>1</sup>, Florian Richter<sup>1</sup>, and Michael C. Yip<sup>1</sup> *Senior Member, IEEE*

**Abstract**—Autonomous suturing has been a long-sought-after goal for surgical robotics. Outside of staged environments, accurate localization of suture needles is a critical foundation for automating various suture needle manipulation tasks in the real world. When localizing a needle held by a gripper, previous work usually tracks them separately without considering their relationship. Because of the significant errors that can arise in the stereo-triangulation of objects and instruments, their reconstructions may often not be consistent. This can lead to unrealistic tool-needle grasp reconstructions that are infeasible. Instead, an obvious strategy to improve localization would be to leverage constraints that arise from contact, thereby constraining reconstructions of objects and instruments into a jointly feasible space. In this work, we consider feasible grasping constraints when tracking the 6D pose of an in-hand suture needle. We propose a reparameterization trick to define a new state space for describing a needle pose, where grasp constraints can be easily defined and satisfied. Our proposed state space and feasible grasping constraints are then incorporated into Bayesian filters for real-time needle localization. In the experiments, we show that our constrained methods outperform previous unconstrained tracking approaches and demonstrate the importance of incorporating feasible grasping constraints into automating suture needle manipulation tasks.

## I. INTRODUCTION

Automating surgical procedures such as suturing has drawn increased interest within the robotics community during the past two decades [1]. The advantage of automation is that it relieves surgeons from time-consuming, tedious, and challenging tasks that often emerge in Minimally Invasive Surgeries [2]–[4]. One of the key components of achieving autonomous procedures is the accurate localization of surgical instruments in the surgical scene [5]–[7]. This localization ability serves as the foundation for automating various surgical tasks in previous work, including needle regrasping [8], [9], knot tying [10], [11], and blood suction [12].

A surgical scene often contains multiple surgical instruments, and previous studies localize them separately without considering their physical interactions [8], [9], [13]–[15]. This can lead to unrealistic environmental reconstruction when combining tracking results of different tools. For example, if a needle is held by a surgical manipulator, tracking them separately can result in the needle being in a non-feasible grasp configuration (e.g., in-collision or floating, as shown in Fig. 1). This can be a dangerous situation because dropping needles can result in damage to surrounding

This research was funded by the Telemedicine and Advanced Technology Research Center (TATRC) as well as NSF Award #2045803.

<sup>1</sup>Zih-Yun Chiu, Florian Richter, and Michael Yip are with the Electrical and Computer Engineering Dept., University of California San Diego, La Jolla, CA 92093 USA. {zchiu, frichter, yip}@ucsd.edu

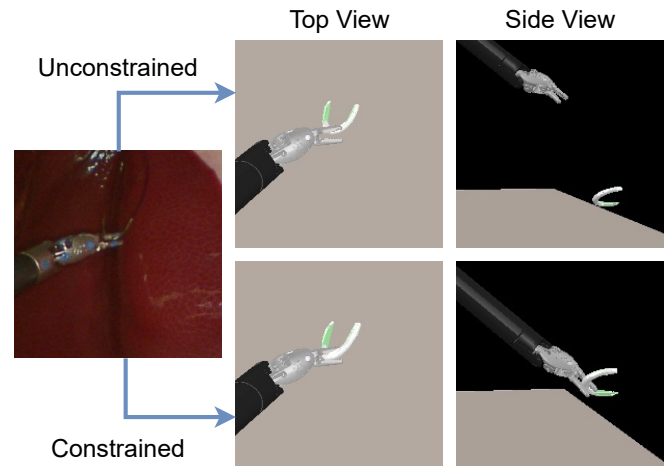


Fig. 1: Live image of a daVinci robot instrument grasping a suture needle, top and side views of tool reconstruction from unconstrained and our constrained needle tracking results. The scene reconstruction of our constrained method is always feasible. This feasibility is not ensured by unconstrained approaches, even when the top view of tool reconstruction aligns well with the live image.

tissue and additional trauma with repetitive needle pick-up. Therefore, in this work, we focus on considering the physical interactions between a suture needle and a surgical manipulator to ensure feasible results in real-time tracking of an in-hand suture needle. Real-time localization is necessary since, in practice, grasping a needle causes it to re-orient in the gripper, and further re-orientation or slippage can happen once the needle interacts with the environment.

### A. Related Work

Current literature on suture needle localization mostly focuses on the features of a needle extracted from camera data and assumes the needle can be anywhere in the space. Several methods reconstruct the pose of a static needle by observing detected markers or learned segmentation [9], [13], [16]. Others have considered uncertainty in the features and motions of a needle and use Bayesian filters with different observation models to track its pose [14], [15], [17]. However, these methods do not consider the physical interactions between the needle and the environment and thus do not guarantee that the needle pose is feasible.

Some studies in suture needle localization consider the physical interactions between a surgical manipulator and a needle when the manipulator holds the needle. One way

is to perform tracking and assume that the configuration between the needle and the manipulator tip is known and remains unchanged over time [18]. Then the robot Jacobian and joint-sensor readings are used to estimate the motions of the needle. However, getting to this known state is nontrivial, as grasping a needle itself is a non-deterministic action, and grasp pose is situation-dependent, such as during regrasping [8]. Thus, the work in [19] does not assume a known configuration of the needle held by an end-effector, and its motions are estimated by a tool-tracking method [7] that tracks the pose of the end-effector. These approaches take into account that the needle should move concurrently with the gripper when held by it. Nonetheless, they do not ensure that the suture needle pose lies inside the feasible grasping manifold of the gripper.

Tracking the poses of an in-hand needle is a constrained pose tracking problem, where the needle should always lie inside the feasible grasping manifold of the gripper. However, there is no unified approach to define a feasible grasping manifold since grippers and grasped objects can be in arbitrary shapes, making this task highly nonlinear. To incorporate constraints into nonlinear tracking problems, previous work follows two approaches: acceptance/rejection sampling [20] and optimization [21], [22]. Acceptance/rejection methods are known to reduce the diversity of the tracked pose [22] and require an excessive number of feasibility checks, making them not desirable for real-time tracking [23]. On the other hand, optimization methods project the estimated pose onto a feasible manifold. However, they require the manifold to be defined as equality or inequality constraints [22], [23], and describing the feasible grasping manifold in such a way would be highly nontrivial.

### B. Contributions

In this work, we achieve state-of-the-art performance for real-time suture needle tracking in robotic surgery by incorporating grasping constraints. To this end, we present the following novel contributions:

- 1) the first approach to probabilistically track a suture needle in real-time with grasping constraints,
- 2) a state-space to describe a grasped suture needle for efficient sampling on the feasible grasping manifold,
- 3) and a comparison of Bayesian filter approaches that incorporate the grasping constraints.

The proposed methods are evaluated in both simulation and real-world environments. In simulation environments, we demonstrate that our proposed methods outperform other unconstrained/constrained tracking approaches. Moreover, we evaluate different tracking methods on the suture needle regrasping task [8], [9]. The results indicate that incorporating grasping constraints makes the regrasping policy more robust to noise in detections. In real-world environments, we use marker-less feature detections from a Deep Neural Network (DNN) as needle observations and reconstruct the tracked tool poses from ex-vivo images. An example is shown in Fig. 1. The results demonstrate that our constrained approach

ensures a feasible estimated pose, and an unconstrained method can lead to unrealistic reconstructions.

## II. METHODS

### A. Problem Formulation

We aim to solve the in-hand suture needle pose,  $\mathbf{s}_t$ , tracking problem probabilistically from a sequence of observations,  $\mathbf{o}_{0:t}$ , which can be formulated as:

$$\begin{aligned} \text{Track } & p_{t|t}(\mathbf{s}_t) := p(\mathbf{s}_t | \mathbf{a}_{0:t-1}, \mathbf{o}_{0:t}) \\ \text{s.t. } & \mathbf{s}_t \in \mathcal{F}_t \\ \text{where } & \mathbf{s}_t = f(\mathbf{s}_{t-1}, \mathbf{a}_{t-1}, \mathbf{w}_{t-1}) \sim p_f(\cdot | \mathbf{s}_{t-1}, \mathbf{a}_{t-1}) \\ & \mathbf{o}_t = h(\mathbf{s}_t, \mathbf{v}_t) \sim p_h(\cdot | \mathbf{s}_t) \end{aligned} \quad (1)$$

where  $\mathcal{F}_t$  is the feasible grasping space,  $f(\cdot)$  and  $h(\cdot)$  are the motion and observation models with noise  $\mathbf{w}_{t-1}$  and  $\mathbf{v}_t$  respectively, and  $\mathbf{a}_{t-1}$  is the action applied to the suture needle.

In our task,  $\mathcal{F}_t$  in (1) is the feasible grasping manifold of the surgical manipulator that is holding a suture needle at time step  $t$ . Usually, a grasping manifold should consider two feasibility constraints: geometric and dynamic constraints. Geometric constraints include [24]:

- 1) The object's surface should be in contact with the gripper's surface, i.e.,  $Surface(\mathbf{s}_t) \cap Surface(\mathbf{e}_t) \neq \emptyset$ , where  $\mathbf{e}_t$  is the state of the gripper at time step  $t$ .
- 2) The object should not penetrate with the gripper, i.e.,  $Interior(\mathbf{s}_t) \cap Interior(\mathbf{e}_t) = \emptyset$ .

Dynamic constraints include that if there is no external force except gravity acting on both the object and the gripper, the linear and angular velocities of the object relative to the gripper should be  $\mathbf{0}$ . Hence, the feasible grasping manifold  $\mathcal{F}_t$  can be represented as

$$\mathcal{F}_t = \{\mathbf{s}_t | \mathbf{s}_t \in \mathcal{G}_t \cap \mathcal{D}_t\}, \quad (2)$$

$$\text{where } \mathcal{G}_t = \{\mathbf{s}_t | Surface(\mathbf{s}_t) \cap Surface(\mathbf{e}_t) \neq \emptyset \text{ and } Interior(\mathbf{s}_t) \cap Interior(\mathbf{e}_t) = \emptyset\}, \quad (3)$$

$$\mathcal{D}_t = \{\mathbf{s}_t | \text{If } ExternalForce \setminus Gravity = \emptyset, \text{ then } LinearVelocity(\mathbf{s}_t, \mathbf{e}_t) = \mathbf{0} \text{ and } AngularVelocity(\mathbf{s}_t, \mathbf{e}_t) = \mathbf{0}\}. \quad (4)$$

Due to the special design and property of surgical manipulators and suture needles, we can simplify the requirements of defining the feasible grasping manifold for an in-hand needle. More specifically, the dynamic constraints in (4) are ignored because (1) a suture needle is very light compared to the gripper, and (2) grippers for surgical manipulators are designed to increase the friction between themselves and the objects they are holding (e.g., Needle Drivers). Hence,  $\mathcal{F}_t = \mathcal{G}_t, \forall t \in [1, \dots, T]$ .

Since the robot end-effector or the grasped object can have a complex shape, the feasible grasping manifold  $\mathcal{F}_t$  in (2) is difficult to define for the object pose,  $[\mathbf{b}_t^\top \ \mathbf{q}_t^\top]^\top$ , where  $\mathbf{b}_t \in \mathbb{R}^3$  is the position, and  $\mathbf{q}_t \in \mathbb{R}^3$  is the axis-angle orientation. The object pose, which is described in a global frame such as the camera frame or in the ego-centric end-effector frame,

is often directly used as the state in (1) [14], [15], [17]–[19]. However, without a feasibility-checking library or a physical simulator, it is difficult to tell whether a pose state  $\mathbf{s}_t^p = [\mathbf{b}_t^\top \mathbf{q}_t^\top]^\top$  belongs to  $\mathcal{F}_t$ . These libraries and simulators require the mesh files of the end-effectors and objects, and multiple proximity queries on geometric models can slow down tracking. Moreover,  $\mathcal{F}_t$  for the pose state is generally irregular, so randomly sampling a pose state from  $\mathcal{F}_t$  can take even more time.

### B. Needle Pose Reparameterization

Instead of defining the state as the needle pose in the camera frame or in the end-effector frame, we propose a **reparameterization trick** that defines a new set of parameters,  $(\alpha, w, u, v)$ , to describe the pose of an in-hand suture needle in the end-effector frame. First, we introduce their *intermediate* parameters,  $(\alpha, d, \theta, \phi)$ , for the pose which was originally defined in our previous work [8]. This set of parameters provides a more intuitive understanding of how to describe the pose of a needle held by a gripper. The first parameter,  $\alpha \in [\frac{1}{2}\pi, \frac{3}{2}\pi] \subseteq \mathbb{R}$ , indicates which point on the needle is grasped, and the other three parameters,  $d \in [d_{min}, d_{max}] \subseteq \mathbb{R}$ ,  $\theta \in [\theta_{min}, \theta_{max}] \subseteq \mathbb{R}$ , and  $\phi \in [\phi_{min}, \phi_{max}] \subseteq \mathbb{R}$ , describe the position of the end-effector relative to the grasped point of the needle in the spherical coordinate system.

Although the  $(\alpha, d, \theta, \phi)$ -parameter is intuitive, independently sampling for  $d, \theta$ , and  $\phi$  can lead to high biases when transforming from spherical coordinates to Cartesian space [25], which is where the suture needle pose is defined. Thus, the following change of variables is applied [26], [27]:

$$w = d^3, \quad u = \frac{\theta}{2\pi}, \quad v = \frac{1}{2}(\cos \phi + 1), \quad (5)$$

where  $w \in [d_{min}^3, d_{max}^3] \subseteq \mathbb{R}$ ,  $u \in [\frac{1}{2\pi}\theta_{min}, \frac{1}{2\pi}\theta_{max}] \subseteq \mathbb{R}$ , and  $v \in [\frac{1}{2}(\cos \phi_{max} + 1), \frac{1}{2}(\cos \phi_{min} + 1)] \subseteq \mathbb{R}$ . Then a reparameterized state  $\mathbf{s}_t^r$  can be represented as  $\mathbf{s}_t^r = [\alpha_t \ w_t \ u_t \ v_t]^\top \in \mathbb{R}^4$ .

Reparameterizing the state of a suture needle as  $(\alpha, w, u, v)$  has several benefits. First, the geometrically feasible space,  $\mathcal{G}_t$  in (3), has a concrete definition:

$$\begin{aligned} \mathcal{G}_t = \mathcal{G} = \{ \mathbf{s}^r = [\alpha \ w \ u \ v]^\top \mid \\ \alpha \in \left[ \frac{\pi}{2}, \frac{3\pi}{2} \right], w \in [d_{min}^3, d_{max}^3], \\ u \in \left[ \frac{\theta_{min}}{2\pi}, \frac{\theta_{max}}{2\pi} \right], \\ v \in \left[ \frac{1}{2}(\cos \phi_{max} + 1), \frac{1}{2}(\cos \phi_{min} + 1) \right] \}. \end{aligned} \quad (6)$$

As long as a state  $\mathbf{s}^r$  belongs to  $\mathcal{G}$ , the needle is in contact with the gripper's surface and not in collision with the end-effector. Hence,  $\mathbf{s}^r$  is a geometrically feasible state. Second, the motion model,  $f(\cdot)$  in (1), for a  $(\alpha, w, u, v)$ -

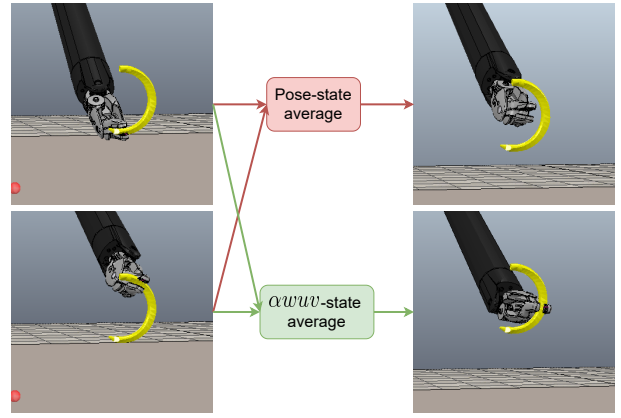


Fig. 2: Average of two pose-states and two  $(\alpha, w, u, v)$ -states. A direct average of two pose-states can lead to an infeasible needle state. However, the direct weighted sum of two  $(\alpha, w, u, v)$ -states is still feasible since the feasible grasping manifold for  $(\alpha, w, u, v)$ -states is convex.

state becomes very simple:

$$\mathbf{s}_{t+1}^r = f(\mathbf{s}_t^r, \mathbf{a}_t^r, \mathbf{w}_t^r) = \text{clip}(\mathbf{s}_t^r + \mathbf{a}_t^r + \mathbf{w}_t^r, \mathbf{s}_{min}^r, \mathbf{s}_{max}^r), \quad (7)$$

$$\mathbf{s}_{min}^r = \left[ \frac{\pi}{2} \quad d_{min}^3 \quad \frac{\theta_{min}}{2\pi} \quad \frac{1}{2}(\cos \phi_{max} + 1) \right]^\top, \quad (8)$$

$$\mathbf{s}_{max}^r = \left[ \frac{3\pi}{2} \quad d_{max}^3 \quad \frac{\theta_{max}}{2\pi} \quad \frac{1}{2}(\cos \phi_{min} + 1) \right]^\top, \quad (9)$$

where  $\mathbf{a}_t^r = [a_{t,\alpha} \ a_{t,w} \ a_{t,u} \ a_{t,v}]^\top \in \mathbb{R}^4$  is the variation applied to  $\mathbf{s}_t^r$ , and  $\mathbf{w}_t^r \in \mathbb{R}^4$  is the motion noise. This simple form ensures feasible outputs from the motion model while requiring no time-consuming post-processing, such as rejection sampling or optimization. Finally, the geometrically feasible space,  $\mathcal{G}$ , when using  $(\alpha, w, u, v)$ -state, is the shape of a hyperrectangle, so it is a convex space. The convexity of  $\mathcal{G}$  enables the direct weighted sum of multiple tracked state candidates to also be feasible, which is important when computing averages. Meanwhile, for pose-states, the feasible space is usually non-convex due to complex gripper and object shapes in the 3D space, and the weighted average of multiple poses can be infeasible. Fig. 2 shows an example of averaging two pose-states and two  $(\alpha, w, u, v)$ -states. The former results in a floating needle, whereas the latter remains feasible grasping.

Mappings between a  $(\alpha, w, u, v)$ -parameter and its corresponding needle pose in the camera frame,  $\mathbf{p}_n^c = [\mathbf{b}_n^{c\top} \mathbf{q}_n^{c\top}]^\top$ , are necessary since the observations of the needle are from the images taken by the camera. The steps to transform from  $(\alpha, w, u, v)$  to  $\mathbf{p}_n^c$  are summarized as follows:

- 1) Transform  $(w, u, v)$  to  $(d, \theta, \phi)$  using the inverse of (5).
- 2) Transform  $(\alpha, d, \theta, \phi)$  to  $\mathbf{p}_n^c$  and  $\mathbf{p}_e^c$  using the methods described in [8], where  $\mathbf{p}_e^c = [\mathbf{b}_e^{n\top} \mathbf{q}_e^{n\top}]^\top$  is the pose of the end-effector in the needle frame.

Transforming from  $\mathbf{p}_n^c$  to  $(\alpha, w, u, v)$  requires some properties of an in-hand suture needle. First, we list the steps of this transformation as follows:

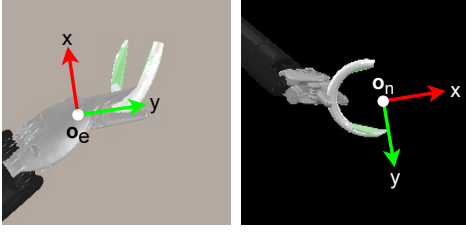


Fig. 3: The top and side views of a surgical manipulator stably grasping a suture needle and the coordinate frames used in this work. It can be observed that when a needle is stably grasped, the y-axis of the end-effector frame will pass through at least one point on the needle.

- 1) Calculate  $\mathbf{p}_e^n$  using  $\mathbf{p}_n^c$  and  $\mathbf{p}_e^c$ , where  $\mathbf{p}_e^c = [\mathbf{b}_e^{c\top} \mathbf{q}_e^{c\top}]^\top$  is the pose of the end-effector in the camera frame.
- 2) Find out the needle's grasped point in the needle frame,  $\mathbf{p}_g^n = [\mathbf{b}_g^{n\top} \mathbf{q}_g^{n\top}]^\top$ .
- 3) Calculate  $(\alpha, d, \theta, \phi)$  using  $\mathbf{p}_g^n$  and  $\mathbf{p}_e^n$ .
- 4) Transform  $(d, \theta, \phi)$  to  $(w, u, v)$  using (5).

The first step can be calculated by

$$\mathbf{H}(\mathbf{b}_e^n, \mathbf{q}_e^n) = (\mathbf{H}(\mathbf{b}_n^c, \mathbf{q}_n^c))^{-1} \mathbf{H}(\mathbf{b}_e^c, \mathbf{q}_e^c). \quad (10)$$

where  $\mathbf{H}(\cdot) \in SE(3)$  is the homogeneous transform from position and orientation vectors. To obtain  $\mathbf{p}_g^n$  in the second step, we use the following property of an in-hand suture needle: *If a needle is stably grasped by a surgical manipulator, the y-axis of the end-effector frame will pass through the needle's grasped point.* An example can be observed in Figure 3. This property allows us to calculate  $\mathbf{p}_g^n$  by

$$\mathbf{b}_g^n = \mathbf{b}_e^n + \beta \mathbf{R}_{e,y}^n, \quad (11)$$

$$\mathbf{q}_g^n = \mathbf{0}, \quad (12)$$

$$\text{where } \beta = -([\mathbf{0} \ 0 \ 1] \mathbf{b}_e^n) / ([\mathbf{0} \ 0 \ 1] \mathbf{R}_{e,y}^n). \quad (13)$$

The coefficient  $\beta \in \mathbb{R}$  in (11) is obtained by *finding the intersection between  $\mathbf{R}_{e,y}^n$  and the xy-plane of the needle frame*, where  $\mathbf{R}_{e,y}^n \in \mathbb{R}^3$  is the y-axis of the end-effector frame in the needle frame. With  $\mathbf{p}_g^n$ ,  $\alpha$  can be calculated by

$$\alpha = \tan^{-1} \left( \frac{b_{g,y}^n}{b_{g,x}^n} \right), \quad (14)$$

where  $b_{g,x}^n \in \mathbb{R}$  and  $b_{g,y}^n \in \mathbb{R}$  are the x- and y-coordinates of  $\mathbf{b}_g^n$  [8]. To obtain  $(d, \theta, \phi)$  in the third step, we first calculate

$$\mathbf{b}_e^g = [b_{e,x}^g \ b_{e,y}^g \ b_{e,z}^g]^\top = \mathbf{b}_e^n - \mathbf{b}_g^n. \quad (15)$$

Then the  $(d, \theta, \phi)$ -parameters become [8]

$$d = \|\mathbf{b}_e^g\|_2, \theta = \tan^{-1} \left( \frac{b_{e,y}^g}{b_{e,x}^g} \right), \phi = \cos^{-1} \left( \frac{b_{e,z}^g}{d} \right). \quad (16)$$

Finally,  $(w, u, v)$  in the fourth step can be calculated by (5).

### C. Bayesian Filter with Feasibility Constraints

Bayesian filtering is applied to solve (1) with the  $(\alpha, w, u, v)$ -state representation, which is easy to sample and constrain on  $\mathcal{G}$ . Due to the nonlinearity in the observation

### Algorithm 1: Constrained Needle Tracking with HF/PF

---

**Input:** maximal and minimal values of  $(\alpha, w, u, v)$ -parameters  $\mathbf{s}_{min}^r$  and  $\mathbf{s}_{max}^r$ , tracking algorithm HF or PF, motion noise covariance  $\Sigma_m^r$ , image observations  $\mathbb{I}_{1:T}$ , end-effector poses  $\mathbf{p}_{e,1:T}^c$ , needle radius  $r$ , observation noise variance  $\sigma_o^2$

**Output:** tracked needle state  $\mathbf{s}_{1:T}^r$

```

// Initialization
1  $\{\mathbf{s}_{0,i}^r\}_{i=1}^n \leftarrow \text{sampleFeasible}(\mathbf{s}_{min}^r, \mathbf{s}_{max}^r)$ 
2  $\{\gamma_{0,i}\}_{i=1}^n \leftarrow \{\frac{1}{n}, \dots, \frac{1}{n}\}$ 
3 for timestep  $t = 1, \dots, T$  do
   // Predict Step
4   if HF then
5      $\{\gamma_{t,i}\}_{i=1}^n \leftarrow \text{Predict}(\{\mathbf{s}_{t-1,i}^r, \gamma_{t-1,i}\}_{i=1}^n, \Sigma_m^r)$ 
6      $\{\mathbf{s}_{t,i}^r\}_{i=1}^n \leftarrow \{\mathbf{s}_{t-1,i}^r\}_{i=1}^n$ 
7   else if PF then
8      $\{\mathbf{s}_{t,i}^r\}_{i=1}^n \leftarrow \text{Predict}(\{\mathbf{s}_{t-1,i}^r\}_{i=1}^n, \Sigma_m^r)$ 
9      $\{\gamma_{t,i}\}_{i=1}^n \leftarrow \{\gamma_{t-1,i}\}_{i=1}^n$ 
10  end
   // Update Step
11   $\mathbf{o}_t \leftarrow \text{getDetections}(\mathbb{I}_t)$ 
12   $\{\mathbf{s}_{t,i}^p\}_{i=1}^n \leftarrow \{\alpha w u v \text{2pose}(\mathbf{s}_{t,i}^r, \mathbf{p}_{e,t}^c, r)\}_{i=1}^n$ 
13   $\{\gamma_{t,i}\}_{i=1}^n \leftarrow \text{weightUpdate}(\{\mathbf{s}_{t,i}^p, \gamma_{t,i}\}_{i=1}^n, \mathbf{o}_t, \sigma_o^2)$ 
14  if PF then
15     $\{\mathbf{s}_{t,i}^r\}_{i=1}^n \leftarrow \{\text{pose2}\alpha w u v(\mathbf{s}_{t,i}^p, \mathbf{p}_{e,t}^c)\}_{i=1}^n$ 
16    // Stratify Resampling [28]
17    if  $\text{effectiveParticles}(\{\gamma_{t,i}\}_{i=1}^n) < N_{eff}$  then
18       $\{\mathbf{s}_{t,i}^r, \gamma_{t,i}\}_{i=1}^n \leftarrow \text{resample}(\{\mathbf{s}_{t,i}^r, \gamma_{t,i}\}_{i=1}^n)$ 
19    end
20  // Return Mean Needle Pose
21   $\mathbf{s}_t^r \leftarrow \sum_{i=1}^n \gamma_{t,i} \mathbf{s}_{t,i}^r$ 

```

---

models, Histogram Filter (HF) and Particle Filter (PF) are both implemented. A summary is provided in Algorithm 1.

*Initialization:* Upon initialization of the filters, a set of discrete states  $\mathcal{S}^r$  such that  $\mathcal{S}^r = \{\mathbf{s}_1^r, \dots, \mathbf{s}_n^r | \mathbf{s}_i^r \in \mathcal{G}, \forall i \in \{1, \dots, n\}\}$  is collected. We sample directly in our proposed  $(\alpha, w, u, v)$ -state space

$$\mathbf{s}^r = [\alpha \ w \ u \ v]^\top, \quad \text{where} \quad (17)$$

$$\alpha \sim \mathcal{U} \left( \frac{\pi}{2}, \frac{3\pi}{2} \right) \quad (18)$$

$$w \sim \mathcal{U} (d_{min}^3, d_{max}^3) \quad (19)$$

$$u \sim \mathcal{U} \left( \frac{\theta_{min}}{2\pi}, \frac{\theta_{max}}{2\pi} \right) \quad (20)$$

$$v \sim \mathcal{U} \left( \frac{1}{2}(\cos \phi_{max} + 1), \frac{1}{2}(\cos \phi_{min} + 1) \right). \quad (21)$$

$\mathcal{U}(\cdot, \cdot)$  is the uniform distribution. The initialization is done in lines 1 and 2 of Algorithm 1.

*Motion Model:* Since our goal is to track a *grasped* suture needle, the motion model's probability distribution is

$$p_f(\cdot | \mathbf{s}_t^r, \mathbf{a}_t^r) \sim \mathcal{N}(\mathbf{s}_t^r + \mathbf{a}_t^r, \Sigma_m^r). \quad (22)$$

where  $\mathbf{a}_t^r = \mathbf{0} \ \forall t \in \{1, \dots, T\}$ , and  $\Sigma_m^r \in \mathbb{R}^{4 \times 4}$  is the covariance matrix of the motion noise. Note that with  $\mathbf{a}_t^r = \mathbf{0}$ , we track the needle poses relative to the surgical manipulator instead of the camera. In addition, this implies that a grasped needle is not expected to move largely inside

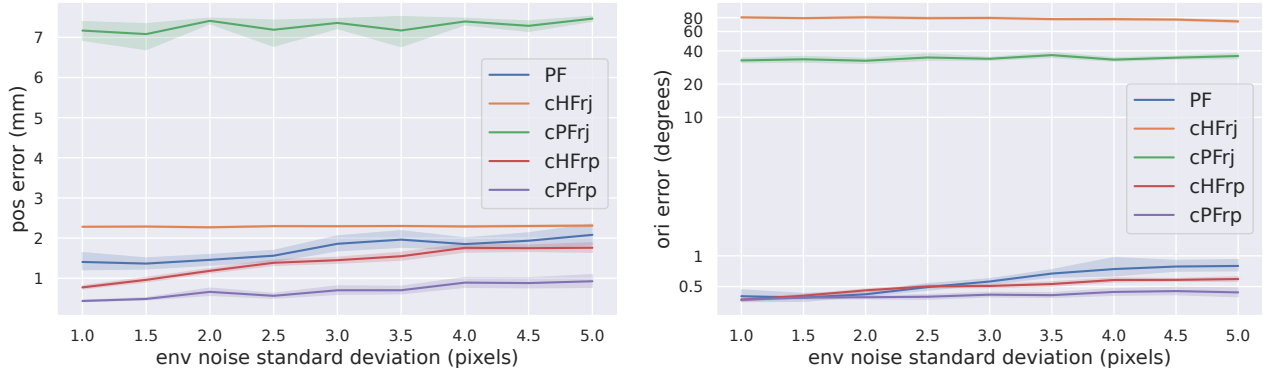


Fig. 4: The positional errors (left) and orientational errors (right, log-scale) of tracking an in-hand suture needle grasped by a moving gripper. Our proposed state-reparameterization methods outperform other approaches since the new state space allows a concrete definition of the feasible grasping manifold and can easily incorporate grasping constraints into tracking.

a gripper. The motion model is applied in lines 4-10 in Algorithm 1.

*Observation Model:* The observation model is our previously proposed *Points Matching to Ellipse Observation Model* [19] which can be applied to any pixel detections of a suture needle. Note that the observation model requires the pose-state,  $\mathbf{s}_t^p = [\mathbf{b}_n^{c\top} \mathbf{q}_n^{c\top}]^\top$ , as the input since the detections are from an image. For more details on the observation model, see [19]. In Algorithm 1, the observation model is applied in lines 11-13, where  $\sigma_o^2 \in \mathbb{R}$  is the variance of the observation model. Additional steps for PF, such as resampling particles, are detailed in Lines 14-19.

*Output:* To get a mean result from the HF and PF algorithms, the weighted average of the discrete states is computed as shown in line 20 in Algorithm 1. Note that  $\mathbf{s}_t^c$  is still feasible since  $\mathcal{G}$  is a convex space.

### III. EXPERIMENT AND RESULTS

#### A. Tracking Experiments in Simulation

We first evaluate our proposed constrained in-hand suture needle pose tracking algorithms in a CoppeliaSim environment. The simulation environmental settings are similar to the ones in [19], where a surgical manipulator holds a radius of 5.4mm suture needle. A stereo camera is set up to capture images with a size of  $256 \times 256$ , and five needle points are extracted from each image as the detections. Each detected point is disturbed by a noise sampled from  $\mathcal{N}(\mathbf{0}, \Sigma_n)$ , where  $\Sigma_n \in \mathbb{R}^{2 \times 2}$  is a diagonal matrix with each diagonal element being  $\sigma_n^2$ . Each experiment is repeated for 20 trials, and each trial contains 100 time steps.

We compare the following state definitions for an in-hand suture needle and tracking algorithms:

- 1) Unconstrained pose-state, PF (PF) [18], [19]: naive PF with pose-states, and the needle motions are set to be the same as that of the end-effector.
- 2) Constrained pose-state by rejection sampling, HF (cHFrfj) [29]: constrained HF with feasible states pre-collected by rejection sampling.
- 3) Constrained pose-state by rejection sampling, PF (cPFrfj) [20]: constrained PF with rejection sampling applied to the outputs of the motion model.
- 4) Reparameterized  $(\alpha, w, u, v)$ -state, HF (cHFrfp): our proposed state space with HF (Algorithm 1).
- 5) Reparameterized  $(\alpha, w, u, v)$ -state, PF (cPFrfp): our proposed state space with PF (Algorithm 1).



For all algorithms, we use the state-of-the-art *Points Matching to Ellipse Observation Model* for needle detections [19]. Each algorithm is run with 2000 particles or pre-collected states. We use the following two equations to calculate the positional and orientational errors of a tracked needle pose:

$$err_{\mathbf{b}} = \|\mathbf{b}_t - \bar{\mathbf{b}}_t\|_2, \quad err_{\mathbf{q}} = \|\text{AxisAngle}(\mathbf{q}_t(\bar{\mathbf{q}}_t)^{-1})\|_2, \quad (23)$$

where  $\bar{\cdot}$  is the ground truth obtained in the simulator, and  $\text{AxisAngle}(\cdot)$  transforms a quaternion to its axis-angle representation.

Fig. 4 shows the pose errors of tracking an in-hand suture needle while moving the end-effector. Although Bayesian filters with rejection sampling consider the feasible grasping constraints, their tracking errors are larger than other methods, even than the unconstrained one. This is because the feasible grasping manifold for pose-states is too irregular and difficult to sample uniformly from. With such an irregular space, the states sampled from it usually lack diversity since some states are easier to be sampled than others. Moreover, it requires lots of samples and feasibility checks to obtain enough feasible states. The average run time of cPFrfj, which runs the rejection sampling process online, is 3.6 seconds per image frame, whereas other methods take 0.15 seconds per image frame.

Tracking an in-hand needle with our reparameterized  $(\alpha, w, u, v)$ -states achieves the lowest pose errors. Sampling from the feasible grasping manifold of  $(\alpha, w, u, v)$ -states is much easier and requires no feasibility checks. Hence, both tracking errors and run time can be much lower than rejection-sampling-based methods. In addition, our proposed methods consider the physical interactions between the gripper and the needle. Therefore, when the environmental noise

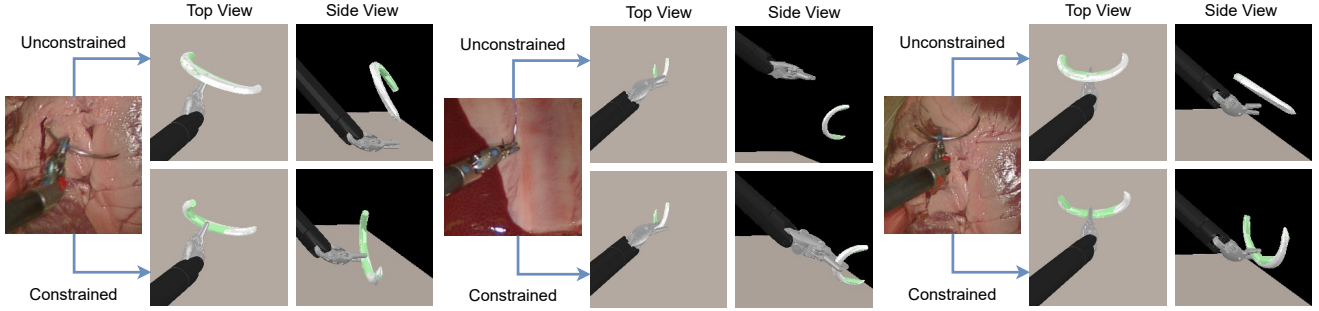


Fig. 5: Raw image, top and side views of tool reconstruction from an unconstrained (PF) and our constrained (cPFrp) needle tracking approaches, across three examples. Without incorporating grasping constraints into tracking, PF often estimates the needle pose with inaccurate depth and orientation, whereas cPFrp provides a more realistic, feasible reconstructed pose.

TABLE I: Success rate for needle regrasping policy [8].

$\sigma_{e,p} = 0$ mm, $\sigma_{e,o} = 0$ degree					
$\sigma_n$ (pixels)	1	2	3	4	5
PF	0.9	0.77	0.6	0.6	0.43
cPFrp	0.97	0.87	0.97	0.87	0.77
$\sigma_{e,p} = 1$ mm, $\sigma_{e,o} = 5$ degrees					
$\sigma_n$ (pixels)	1	2	3	4	5
PF	0.67	0.6	0.6	0.37	0.4
cPFrp	0.83	0.9	0.9	0.87	0.87
$\sigma_{e,p} = 2$ mm, $\sigma_{e,o} = 10$ degrees					
$\sigma_n$ (pixels)	1	2	3	4	5
PF	0	0.07	0.03	0	0
cPFrp	0.63	0.67	0.6	0.57	0.73

increases, the tracking errors of our methods remain low compared to the unconstrained method. Finally, the performance of cPFrp is better than cHFrp since PF allows online adjustment of the tracked state candidates. This adjustment moves those candidates closer to the real state.

### B. Automated Suture Needle Regrasping

To demonstrate the importance of integrating grasping constraints into suture needle tracking, we compare the unconstrained PF and our proposed cPFrp methods on the suture needle regrasping task [8] in the simulation environment. In these experiments, we add noise not only to needle detections but also to the detected end-effector poses. The positional noise of end-effector poses is sampled from  $\mathcal{N}(\mathbf{0}, \Sigma_{e,p})$ , where  $\Sigma_{e,p} \in \mathbb{R}^{3 \times 3}$  is a diagonal matrix with each diagonal element being  $\sigma_{e,p}^2$ . The orientational noise of the end-effector pose is  $\theta_e [0 \ 1 \ 0]^T$ , where  $\theta_e \in \mathbb{R}$  is sampled from  $\mathcal{N}(0, \sigma_{e,o}^2)$ . Each experiment is run for 30 trials.

Table I shows the regrasping success rate of PF and cPFrp under different environmental noise. PF barely shows any success when more noise is added to detected end-effector poses. On the other hand, cPFrp succeeds in multiple regrasping trials under the same condition. These results suggest that when automating suture needle manipulation tasks, it is essential to consider the relationship between the needle and the gripper that is interacting with it.

### C. Tracking Experiments in Real World

We evaluate PF and our cPFrp methods on ex-vivo datasets and demonstrate the tool reconstruction results. The suture

needles tested are with radii 7mm and 11.5mm, and the algorithms are run with 2000 particles. In the ex-vivo datasets, a needle is grasped by a Large Needle Driver (LND) installed on a Patient Side Manipulator (PSM) arm of the da Vinci Research Kit (dVRK) [30]. dVRK's stereo endoscopic camera, which is 1080p and runs with 30 fps, is used to capture images for end-effector and needle detections. The end-effector poses are tracked by our previous method [7], and the markerless needle detections are obtained by DeepLabCut [31], the state-of-the-art keypoint detector.

Figure 1 and 5 show the top and side views of tool reconstruction. From the top views, the results of both the unconstrained and constrained methods align well with the raw images. However, the side views clearly show their differences. The unconstrained method can lead to the reconstructed needle floating in the air or in collision with the gripper. Also, the inaccuracy mainly happens when reconstructing the depths of a needle from the camera. This is because the detections of a needle with different depths do not show many differences in the endoscopic images. Therefore, without integrating the feasibility constraints into needle tracking, the unconstrained method is likely to estimate a needle pose with inaccurate depth.

## IV. DISCUSSION AND CONCLUSION

In this work, we propose approaches to incorporate feasible grasping constraints into real-time localizing the 6D pose of an in-hand suture needle. We define a novel state space for the 6D pose of a needle. This state space allows efficient sampling from the feasible grasping manifold, which requires no feasibility checks with external software. We incorporate the proposed state space and feasible grasping constraints into Bayesian filters for real-time needle pose tracking and demonstrate that our constrained approaches outperform previous ones.

Our proposed constrained method ensures the needle pose is always feasible given the estimated end-effector pose. Note that although there can be noise in end-effector poses, the reconstructed relative pose between the needle and the end-effector is accurate for our constrained method. In Table I, we show that the accuracy of this relative pose is the key to success in automating suture needle manipulation tasks.

## REFERENCES

- [1] M. Yip and N. Das, "Robot autonomy for surgery," in *The Encyclopedia of MEDICAL ROBOTICS: Volume 1 Minimally Invasive Surgical Robotics*. World Scientific, 2019, pp. 281–313.
- [2] A. Garcia-Ruiz, M. Gagner, J. H. Miller, C. P. Steiner, and J. F. Hahn, "Manual vs robotically assisted laparoscopic surgery in the performance of basic manipulation and suturing tasks," *Archives of surgery*, vol. 133, no. 9, pp. 957–961, 1998.
- [3] G. Hubens, H. Coveliers, L. Balliu, M. Ruppert, and W. Vaneerdeweg, "A performance study comparing manual and robotically assisted laparoscopic surgery using the da vinci system," *Surgical Endoscopy and other interventional techniques*, vol. 17, no. 10, pp. 1595–1599, 2003.
- [4] F. Corcione, C. Esposito, D. Cucurullo, A. Settembre, N. Miranda, F. Amato, F. Pirozzi, and P. Caiazzo, "Advantages and limits of robot-assisted laparoscopic surgery: preliminary experience," *Surgical Endoscopy and Other Interventional Techniques*, vol. 19, no. 1, pp. 117–119, 2005.
- [5] Y. Li, F. Richter, J. Lu, E. K. Funk, R. K. Orosco, J. Zhu, and M. C. Yip, "Super: A surgical perception framework for endoscopic tissue manipulation with surgical robotics," *IEEE Robotics and Automation Letters*, vol. 5, no. 2, pp. 2294–2301, 2020.
- [6] J. Lu, A. Jayakumari, F. Richter, Y. Li, and M. C. Yip, "Super deep: A surgical perception framework for robotic tissue manipulation using deep learning for feature extraction," in *2021 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2021, pp. 4783–4789.
- [7] F. Richter, J. Lu, R. K. Orosco, and M. C. Yip, "Robotic tool tracking under partially visible kinematic chain: A unified approach," *IEEE Transactions on Robotics*, 2021.
- [8] Z.-Y. Chiu, F. Richter, E. K. Funk, R. K. Orosco, and M. C. Yip, "Bimanual regrasping for suture needles using reinforcement learning for rapid motion planning," in *2021 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2021, pp. 7737–7743.
- [9] A. Wilcox, J. Kerr, B. Thananjeyan, J. Ichnowski, M. Hwang, S. Paradis, D. Fer, and K. Goldberg, "Learning to localize, grasp, and hand over unmodified surgical needles," in *2022 International Conference on Robotics and Automation (ICRA)*. IEEE, 2022, pp. 9637–9643.
- [10] D.-L. Chow and W. Newman, "Improved knot-tying methods for autonomous robot surgery," in *2013 IEEE International Conference on Automation Science and Engineering (CASE)*. IEEE, 2013, pp. 461–465.
- [11] B. Lu, H. K. Chu, K. Huang, and L. Cheng, "Vision-based surgical suture looping through trajectory planning for wound suturing," *IEEE Transactions on Automation Science and Engineering*, vol. 16, no. 2, pp. 542–556, 2018.
- [12] F. Richter, S. Shen, F. Liu, J. Huang, E. K. Funk, R. K. Orosco, and M. C. Yip, "Autonomous robotic suction to clear the surgical field for hemostasis using image-based blood flow detection," *IEEE Robotics and Automation Letters*, vol. 6, no. 2, pp. 1383–1390, 2021.
- [13] S. Iyer, T. Looi, and J. Drake, "A single arm, single camera system for automated suturing," in *2013 IEEE International Conference on Robotics and Automation*. IEEE, 2013, pp. 239–244.
- [14] M. Ferro, G. Fontanelli, F. Ficuciello, B. Siciliano, and M. Vendittelli, "Vision-based suturing needle tracking with extended kalman filter," in *Computer/Robot Assisted Surgery workshop*, 2017.
- [15] 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 international conference on robotics and automation (ICRA)*. IEEE, 2016, pp. 4178–4185.
- [16] 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 International Conference on Robotics and Automation (ICRA)*. IEEE, 2018, pp. 1370–1377.
- [17] Y. Kurose, Y. M. Baek, Y. Kamei, S. Tanaka, K. Harada, S. Sora, A. Morita, N. Sugita, and M. Mitsuishi, "Preliminary study of needle tracking in a microsurgical robotic system for automated operations," in *2013 13th international conference on control, automation and systems (ICCAS 2013)*. IEEE, 2013, pp. 627–630.
- [18] O. Özgüner, R. Hao, R. C. Jackson, T. Shkurti, W. Newman, and M. C. Cavusoglu, "Three-dimensional surgical needle localization and tracking using stereo endoscopic image streams," in *2018 IEEE international conference on robotics and automation (ICRA)*. IEEE, 2018, pp. 6617–6624.
- [19] Z.-Y. Chiu, A. Z. Liao, F. Richter, B. Johnson, and M. C. Yip, "Markerless suture needle 6d pose tracking with robust uncertainty estimation for autonomous minimally invasive robotic surgery," *arXiv preprint arXiv:2109.12722*, 2021.
- [20] L. Lang, W.-s. Chen, B. R. Bakshi, P. K. Goel, and S. Ungarala, "Bayesian estimation via sequential monte carlo sampling—constrained dynamic systems," *Automatica*, vol. 43, no. 9, pp. 1615–1622, 2007.
- [21] A. Kong, J. S. Liu, and W. H. Wong, "Sequential imputations and bayesian missing data problems," *Journal of the American statistical association*, vol. 89, no. 425, pp. 278–288, 1994.
- [22] Z. Zhao, B. Huang, and F. Liu, "Constrained particle filtering methods for state estimation of nonlinear process," *AICHE Journal*, vol. 60, no. 6, pp. 2072–2082, 2014.
- [23] C. Hu, Y. Liang, X. Wang, and L. Xu, "A particle filter via constrained sampling for nonlinear dynamic systems," *International Journal of Robust and Nonlinear Control*, vol. 30, no. 13, pp. 4944–4959, 2020.
- [24] Y. Hasson, G. Varol, D. Tzionas, I. Kalevatykh, M. J. Black, I. Laptev, and C. Schmid, "Learning joint reconstruction of hands and manipulated objects," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2019, pp. 11 807–11 816.
- [25] E. W. Weisstein, "Sphere point picking," <https://mathworld.wolfram.com/>, 2002.
- [26] H. Cundy and A. Rollett, "Sphere and cylinder—archimedes' theorem," *Mathematical Models*, pp. 172–173, 1989.
- [27] R. Harman and V. Lacko, "On decompositional algorithms for uniform sampling from n-spheres and n-balls," *Journal of Multivariate Analysis*, vol. 101, no. 10, pp. 2297–2304, 2010.
- [28] G. Kitagawa, "Monte carlo filter and smoother for non-gaussian nonlinear state space models," *Journal of computational and graphical statistics*, vol. 5, no. 1, pp. 1–25, 1996.
- [29] N. Boyko and Y. Hladun, "Histogram filter for robot localization," in *2021 IEEE 16th International Conference on Computer Sciences and Information Technologies (CSIT)*, vol. 1. IEEE, 2021, pp. 38–43.
- [30] 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 international conference on robotics and automation (ICRA)*. IEEE, 2014, pp. 6434–6439.
- [31] A. Mathis, P. Mamidanna, K. M. Cury, T. Abe, V. N. Murthy, M. W. Mathis, and M. Bethge, "DeepLabcut: markerless pose estimation of user-defined body parts with deep learning," *Nature neuroscience*, vol. 21, no. 9, pp. 1281–1289, 2018.