

# Proprioceptive State Estimation for Amphibious Tactile Sensing

Ning Guo , *Student Member, IEEE*, Xudong Han , *Student Member, IEEE*, Shuqiao Zhong , Zhiyuan Zhou , Jian Lin , Jian S. Dai , *Fellow, IEEE*, Fang Wan , *Member, IEEE*, and Chaoyang Song , *Senior Member, IEEE*

**Abstract**—This article presents a novel vision-based proprioception approach for a soft robotic finger that can estimate and reconstruct tactile interactions in terrestrial and aquatic environments. The key to this system lies in the finger’s unique metamaterial structure, which facilitates omnidirectional passive adaptation during grasping, protecting delicate objects across diverse scenarios. A compact in-finger camera captures high-framerate images of the finger’s deformation during contact, extracting crucial tactile data in real time. We present a volumetric discretized model of the soft finger and use the geometry constraints captured by the camera to find the optimal estimation of the deformed shape. The approach is benchmarked using a motion capture system with sparse markers and a haptic device with dense measurements. Both results show state-of-the-art accuracy, with a median error of 1.96 mm for overall body deformation, corresponding to 2.1% of the finger’s length. More importantly, the state estimation is robust in both on-land and underwater environments as we demonstrate its usage for underwater object shape sensing. This combination of passive adaptation and real-time tactile sensing paves the way for amphibious robotic grasping applications.

**Index Terms**—Proprioception, shape reconstruction, soft robotics, state estimation, vision-based tactile sensing (VBTS).

## I. INTRODUCTION

**P**ROPRIOCEPTIVE state estimation (PropSE) refers to the process of determining the internal state or position of

a robot or a robotic component (such as a limb or joint) by measuring the robot’s internal properties [1], [2]. PropSE is particularly important in soft robotics, especially in terrestrial and aquatic environments, where these robots’ flexible and deformable nature makes traditional position and orientation sensing challenging [3]. During the robot’s physical exchange with the external environment, the moment of touch holds the truth of the dynamic interactions [4]. For most living organisms, the skin is crucial in translating material properties, object physics, and interactive dynamics via the sensory receptors into chemical signals [5]. When processed by the brain, they collectively formulate a *feeling* of the external environment (exteroception) [6] and the bodily self (proprioception) [7]. Toward tactile robotics, one stream of research aims at replicating the skin’s basic functionality with comparable or superior performances [8]. For example, developing novel tactile sensors [9] represents a significant research focus. Another stream of research considers robots while developing or utilizing tactile sensors [10]. It requires an interdisciplinary approach to resolve the design challenge involved [11], fostering a growing interest in tactile robotics among academia and industry [12].

We previously conducted a preliminary investigation on vision-based tactile sensing (VBTS) [13], which leverages the visual features of a series of soft metamaterial structures’ large-scale, omnidirectional adaptative deformation. The design of these metamaterial structures was subsequently generalized as a class of soft polyhedral networks (SPNs) [14], for which high-performance proprioceptive learning in object manipulation was achieved via a node-based representation. Recent literature shows the growing adoption of volumetric representation with finite element modeling as the de facto ground truth for soft, dynamic interactions [15]. Yet, the high computational cost limits its application in robotic tasks, where real-time perception is critical [16]. Aquatic machine vision remains difficult [17] for unstructured underwater exploration with changing turbidity [relative clarity of a liquid measured by nephelometric turbidity unit (NTU)]. Finger-based PropSE complements aquatic machine vision by providing localized tactile perception in simultaneous localization and mapping (SLAM) [18]. It is a research gap to investigate the design and learning tradeoff between high-fidelity PropSE and real-time perception in an amphibious environment [3], [15], [19]. In such scenarios, in-finger vision with soft robotic fingers may provide a promising solution to advance the field of tactile robotics.

Received 6 April 2024; revised 18 July 2024; accepted 4 September 2024. Date of publication 18 September 2024; date of current version 17 October 2024. This work was supported in part by the National Natural Science Foundation of China under Grant 62206119 and Grant 62473189, in part by the Science, Technology, and Innovation Commission of Shenzhen Municipality under Grant JCYJ20220818100417038 and Grant JSGG20220831110002004, in part by the Shenzhen Long-Term Support for Higher Education at SUSTech under Grant 20231115141649002, and in part by the SUSTech Virtual Teaching Lab for Machine Intelligence Design and Learning under Grant XJZLGC202241. This article was recommended for publication by Associate Editor H. Zhao and Editor J. Bohg upon evaluation of the reviewers’ comments. (Ning Guo, Xudong Han, and Shuqiao Zhong are co-first authors.) (Corresponding authors: Fang Wan; Chaoyang Song.)

Ning Guo, Xudong Han, Jian S. Dai, and Chaoyang Song are with the Department of Mechanical and Energy Engineering, Southern University of Science and Technology, Shenzhen 518055, China (e-mail: 11930729@mail.sustech.edu.cn; 12231112@mail.sustech.edu.cn; daijs@sustech.edu.cn; songcy@iee.org).

Shuqiao Zhong, Zhiyuan Zhou, and Jian Lin are with the Department of Ocean Science and Engineering, Southern University of Science and Technology, Shenzhen 518055, China (e-mail: 12231088@mail.sustech.edu.cn; zhouzy@sustech.edu.cn; linj@sustech.edu.cn).

Fang Wan is with the School of Design, Southern University of Science and Technology, Shenzhen 518055, China (e-mail: wanf@sustech.edu.cn).

This article has supplementary downloadable material available at <https://doi.org/10.1109/TRO.2024.3463509>, provided by the authors.

Digital Object Identifier 10.1109/TRO.2024.3463509

This article introduces a VBTS approach for real-time and high-fidelity PropSE with demonstrated amphibious applications in the lab and field. This is achieved using the SPN structure with marker-based in-finger vision as the soft robotic fingers for large-scale, omnidirectional adaptations with amphibious tactile sensing capability. We proposed a model-based approach for PropSE by introducing rigidity-aware aggregated multihandle (AMH) constraints to optimize a volumetric parameterization of the soft robotic finger's morphological deformation. This enabled us to restructure the VBTS problem as an implicit surface model using Gaussian processes (GPs) for object shape reconstruction. We benchmarked our proposed method in shape reconstruction against existing solutions with verified superior performances. We also conducted experiments using commercial-grade motion-capture systems and touch-haptic devices, demonstrating our solution's large-scale reconstruction and touch-point estimation performances. Finally, we demonstrated the application of our proposed solutions for amphibious tactile sensing in three experiments, including a shape reconstruction experiment, a turbidity benchmarking experiment, and a tactile grasping experiment on an underwater remotely operated vehicle (ROV). The following are the contributions of this study.

- 1) Modeled PropSE via rigidity-aware AMH constraints.
- 2) Formulated VBTS via an implicit surface model for object shape reconstruction.
- 3) Achieved PropSE for VBTS using SPNs with in-finger vision as robotic tactile fingertips.
- 4) Benchmarked PropSE for amphibious tactile reconstruction with demonstrated applications and testing.

The rest of this article is organized as follows. Section II briefly reviews related literature about the role of PropSE in tactile robotics and its application in amphibious tactile sensing. Section III introduces the soft robotic fingertips for this study and presents our proposed model for PropSE via rigidity-aware aggregated multihandle constraints. This section also formulates our proposed VBTS method via implicit surface modeling. All experimental results are presented in Section IV, including those for benchmarking our proposed method's performance and those conducted explicitly for amphibious tactile sensing underwater. Finally, Section V concludes this article.

## II. LITERATURE REVIEW

### A. Toward Dense Sensing for Tactile Robotics

Tactile sensory generally involves many properties that can be digitized for robotics [20]. For mechanics-based dynamics and control, the interactive forces and torques on the contact surface are a primary concern in robotics [21]. It usually involves a certain level of material softness or structural deformation for an enhanced representation of the mechanic interactions as tactile data. The following are the three general research streams in this field.

1) *Pointwise Sensing in 6-D FT*: Estimating forces at contact points is paramount in robotic systems, enabling awareness of physical interaction between the robot and its surrounding objects [22]. Robotic research, especially when dynamics and mechanics are involved, is generally more interested in utilizing the

force-and-torque (FT) properties for manipulation problems by robotic hands [23] or locomotion tasks by legged systems [24]. The FT properties could be succinctly represented by a 6-D vector of forces and torques for a single reference point, making it comparable to the joint torque sensing in articulated robotic structures. However, the shortcut between physical contact and a pointwise 6-D FT measurement may not capture the full extent of contact information for further algorithmic processing [25].

2) *Bioinspired Sparse Sensing Array*: Similar to the biological skin's super-resolutive mechanoreception for tactile sensing [26], a common approach in engineering is to place an array of sensing units on the interactive surface [27]. Instead of going for a localized 6-D force and torque contact information, researchers usually tackle the problem with enhanced pressure sensing across its entire surface from spatially distributed sensing elements [28]. As a result, one can build models or implement learning algorithms to achieve superresolution by sampling the discrete sensory inputs. This approach continuously estimates the tactile interaction on the surface at a much higher resolution than the sensing array arrangement. Yan et al. [29] showed that one can leverage magnetic properties to achieve decoupled normal and shear forces with simultaneous superresolution in tactile sensing of the normal and frictional forces for high-performing grasping.

3) *Visuo-Tactile Dense Image Sensing*: VBTS recently emerged as a popular approach to significantly increase the sensing resolution [30]. This approach leverages the modern imaging process to visually track the deformation of a soft medium as the interface of physical interaction [31], [32], eliminating the need for biologically inspired superresolution [33]. Robotic vision has already become a primary sensing modality for advanced robots [34]. The maturity of modern imaging technologies drives the hardware to be more compact while the software is more accessible to various algorithm libraries for real-time processing. While the high resolution of modern cameras offers significant advantages, the infinite number of potential configurations of the soft medium introduces a considerable challenge [35].

### B. Proprioceptive State Estimation

For tactile applications in robotics, proprioceptive perception of joint position and body movement plays a critical role in achieving state estimation. The tactile interface is a physical separation between the intrinsic proprioception concerning the robot and the extrinsic perception concerning the object-centric environment. We focus on vision-based proprioception, which also applies to analyzing the abovementioned methods.

1) *Intrinsic Proprioception in Tactile Robotics*: For vision-based intrinsic proprioception, the analysis is usually centered on estimating the state of the soft medium during contact, inferring tactile interaction [36]. To establish a physical correspondence between a finite parameterization state estimation model and an infinite configuration of soft deformation [37], markers that are easy to track are often used to discretize the displacement field of soft mediums. Yamaguchi and Atkeson [38] introduced a simple blob detection method to track uniformly distributed markers in a planar transparent soft layer for deformation approximation. Advanced image analysis [39] is also adopted to

utilize machine learning algorithms to extract high-level deformation patterns from randomly spread markers over the entire 3-D volume of soft medium for robust state estimation [40]. Zhang et al. [41] showed a promising approach to integrate physics-based models that capture the dynamic behavior of the soft medium under deformation.

2) *Extrinsic Perception for Tactile Robotics*: For extrinsic perception, the focus is shifted to estimating the object-level information. Tactile sensing data such as object localization, shape, and dynamics parameters could be used for task-based manipulation and locomotion [20]. Using contact to estimate an object's global geometry is instrumental for intelligent agents to make better decisions during object manipulation [42]. Usually, tactile sensing is employed for estimating the object's shape in visually occluded regions, thus playing a complementary role to vision sensors [43], [44]. However, in scenarios where a structured environment with reliable external cameras is unavailable or impractical, such as during exploration tasks in unstructured environments, tactile sensing can provide valuable feedback to achieve environmental awareness [45].

### C. Amphibious Tactile Robotics

Amphibious environments present a unique and dynamic challenge for robotic systems [46]. Robots operating in these environments must contend with vastly different physical properties, including changes in buoyancy, friction, and fluid dynamics [47]. Furthermore, the transition between water and air requires robots to adapt their sensory systems and control strategies to function effectively in each medium [48].

Developing effective tactile sensors for amphibious robots presents several challenges. Sensors must be robust enough to withstand the harsh aquatic environment and be sensitive enough to detect subtle changes in water and air [49]. The transition between these two media can also cause sensor drift and require calibration to maintain accuracy [50]. Despite these challenges, there are exciting opportunities in amphibious tactile robotics, with improved sensitivity, durability, and resistance to environmental factors [51]. However, a research gap remains in developing an effective tactile sensing method with an integrated finger-based design that directly applies to amphibious applications.

## III. MATERIALS AND METHODS

### A. SPN With In-Finger Vision

Soft grippers can achieve diverse and robust grasping behaviors with a relatively simple control strategy [52]. In this study, we adopted our previous work in a class of SPNs with in-finger vision as the soft robotic finger [13], [14]. As shown in Fig. 1(a), the specific design is modified using an enhanced mounting plate to fix the soft finger and made waterproof for amphibious tactile sensing. The soft finger features a shrinking cross-sectional network design toward the tip, capable of omnidirectional adaptation during physical interactions, as shown in Fig. 1(b). We fabricated the finger by vacuum molding using Hei-cast 8400, a three-component polyurethane elastomer. Based on our previous

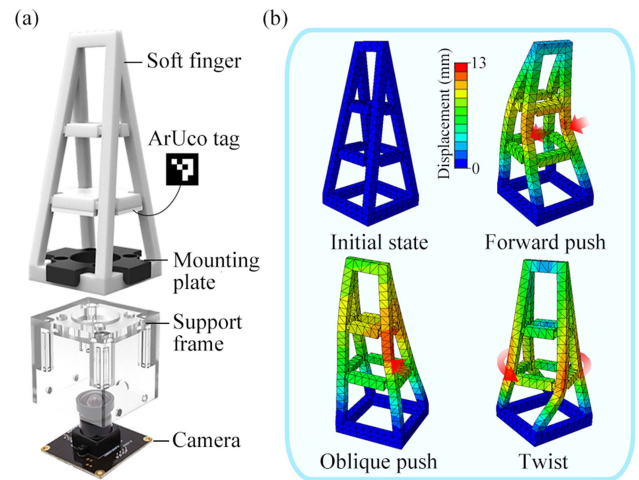


Fig. 1. Assembly and omnidaptive capability of the soft finger. (a) Assembly consists of a soft finger, a rigid plate pasted with an ArUco tag, a mounting plate, a support frame, and a camera. (b) Finger deformation by forward push, oblique push, and twist shows the omnidaptive capability.

work, we mixed the three components with a ratio of 1:1:0, producing a hardness of 90 (Type A) to achieve reliable spatial adaptation for grasping.

An ArUco<sup>1</sup> tag [53] is attached to the bottom side of a rigid plate mechanically fixed with the four lower crossbeams of the soft finger. A monocular RGB camera with a field of view (FOV) of 130° is fixed at the bottom inside a transparent support frame as in-finger vision, video-recording in a high frame rate of 120 frames per second (FPS) at 640 × 480 pixels resolution. When the soft robotic finger interacts with the external environment, live video streams captured by the in-finger vision camera provide real-time pose data of the ArUco tag as rigid-soft kinematics coupling constraints for the PropSE of the soft robotic finger. This marker-based in-finger vision design is equivalent to a miniature motion capture system, efficiently converting the soft robotic finger's spatial deformation into real-time 6-D pose data.

### B. Volumetric Modeling of Soft Deformation for PropSE

Our proposed solution begins by formulating a volumetric model of the soft robotic finger in a 3-D space  $\Omega \in \mathbb{R}^3$  filled with homogeneous elastic material. The distribution of the internal elastic energy within the volumetric elements varies significantly depending on the boundary conditions defined. The PropSE process requires an accurate determination of a smooth deformation map,  $\Phi : \Omega \rightarrow \tilde{\Omega}$ , that facilitates the geometric transformation of the soft body from its initial state, represented by  $\Omega$ , to a deformed state, denoted as  $\tilde{\Omega}$ . This transformation is characterized by minimizing a form of variational energy measuring the distortion of the soft body [54]. As a result, the PropSE performance depends on finite element discretization and the choice of energy function that characterizes deformation.

1) *Volumetric Parameterization of Whole-Body Deformation*: We denote a tetrahedral mesh of the discretized soft

<sup>1</sup><http://sourceforge.net/projects/aruco/>



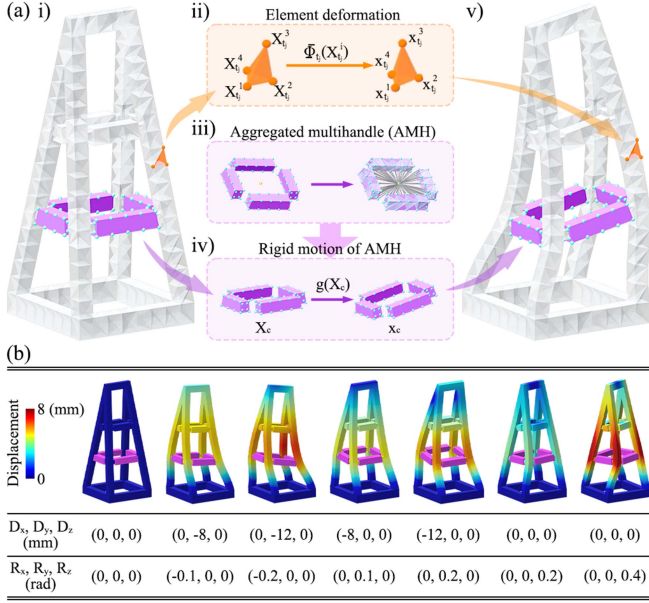


Fig. 2. Proprioceptive deformation modeling and estimation of omnidaptive soft finger. (a) Representation of the proprioceptive model, including i) initial undeformed configuration  $\Omega$  of the soft finger, discretized using tetrahedral mesh; ii) local affine mapping  $\Phi_{t_j}$  applies on  $t_j$  element, transforming each vertex from  $\mathbf{X}_{t_j}^i \in \mathbb{R}^3$  to  $\mathbf{x}_{t_j}^i \in \mathbb{R}^3$ ,  $i \in \{1, 2, 3, 4\}$ ; iii) approximation of visual observed marker area as AMH on the tetrahedral mesh (purple); and iv) applies uniform rigid motion  $g \in SE(3)$  on all AMH that drives soft finger to a deformed configuration  $\tilde{\Omega}$ . (b) Demonstration of soft finger deformation reconstructions under a series of rigid motions applied on AMH, including bending and twisting.

body using  $\mathcal{M} = \{\mathcal{V}, \mathcal{T}\}$ , where  $\mathcal{V} = \{\mathbf{x}_1, \dots, \mathbf{x}_n\}$  is the set of vertices  $\mathbf{x}_i \in \mathbb{R}^3$ , and  $\mathcal{T} = \{t_1, \dots, t_m\}$  is the set of tetrahedra elements, as shown in Fig. 2(a)(i).

When the soft body deforms, a collection of chosen linearly approximated local deformation maps are applied to  $\mathcal{M}$  over each tetrahedron element  $t_j$  via an affine transformation

$$\Phi|_{t_j}(\mathbf{X}) = \mathbf{A}_{t_j} \mathbf{X} + \mathbf{b}_{t_j} \quad (1)$$

where  $\mathbf{X} \in \mathbb{R}^3$  stands for all points inside element  $t_j$ ,  $\mathbf{A}_{t_j} \in \mathbb{R}^{3 \times 3}$  is the differential part of the deformation map, and  $\mathbf{b}_{t_j} \in \mathbb{R}^3$  is the translational part. We choose this piecewise linear deformation map for computational efficiency. High-order deformation functions can be used for better approximation if needed [55].

As shown in Fig. 2(a)(ii), for any  $t_j$  element, the local affine transformation applied on each vertex is denoted as

$$[\mathbf{A}_{t_j} \ \mathbf{b}_{t_j}] \begin{bmatrix} \mathbf{X}_{t_j}^1 & \mathbf{X}_{t_j}^2 & \mathbf{X}_{t_j}^3 & \mathbf{X}_{t_j}^4 \\ 1 & 1 & 1 & 1 \end{bmatrix} = [\mathbf{x}_{t_j}^1 \ \mathbf{x}_{t_j}^2 \ \mathbf{x}_{t_j}^3 \ \mathbf{x}_{t_j}^4] \quad (2)$$

where  $\mathbf{x}_{t_j}^i \in \mathbb{R}^3$ ,  $i \in \{1, 2, 3, 4\}$  are the deformed vertices location of  $t_j$  tetrahedron, and  $\mathbf{X}_{t_j}^i \in \mathbb{R}^3$  are the corresponding initial vertices location.

Therefore, the deformation gradient  $\mathbf{A}_{t_j}$  in the chosen piecewise linear transformation in (1) can be expressed as a linear combination of unknown deformed element vertices location

$\mathbf{x}_{t_j}$  using the following formulation:

$$\mathbf{A}_{t_j}(\mathbf{x}_{t_j}) = \frac{\partial \Phi|_{t_j}}{\partial \mathbf{X}} = \mathbf{D}_s(\mathbf{x}_{t_j}) \cdot \mathbf{D}_m^{-1}(\mathbf{X}_{t_j}) \quad (3)$$

where

$$\mathbf{D}_s(\mathbf{x}_{t_j}) = [\mathbf{x}_{t_j}^2 - \mathbf{x}_{t_j}^1 \quad \mathbf{x}_{t_j}^3 - \mathbf{x}_{t_j}^1 \quad \mathbf{x}_{t_j}^4 - \mathbf{x}_{t_j}^1] \quad (4)$$

$$\mathbf{D}_m(\mathbf{X}_{t_j}) = [\mathbf{X}_{t_j}^2 - \mathbf{X}_{t_j}^1 \quad \mathbf{X}_{t_j}^3 - \mathbf{X}_{t_j}^1 \quad \mathbf{X}_{t_j}^4 - \mathbf{X}_{t_j}^1]. \quad (5)$$

For a discretized tetrahedral mesh  $\mathcal{M}$ , the collection of deformation maps  $\{\Phi_{t_j}\}_{t_j \in \mathcal{T}}$  for all tetrahedra elements should uniquely determine the deformed shape of the soft body [56].

2) *Geometry-Related Deformation Energy Function*: To mimic the physical deformation behavior, the specific energy function form of the deformation map  $\Psi(\Phi_{t_j})$  needs to be specified. Several formulations of geometry-related deformation energies, such as as-rigid-as-possible (ARAP) [57], conformal distortion [58], and isometric distortion [59], have been proposed in recent literature.

Instead of deriving the energy of the system explicitly using constitutive relation and balance equations [60], we choose a symmetric Dirichlet form of energy function [61] to characterize the deformation, which indicates isometric distortion and behaves well in the case of our soft finger. Since the deformation should be irrelevant to the translation, the discrete element energy function only takes the gradient augment of each deformation maps  $\{\Phi_{t_j}\}_{t_j \in \mathcal{T}}$  as

$$\Psi(\Phi_{t_j}) = \Psi(\mathbf{A}_{t_j}) = \|\mathbf{A}_{t_j}\|_{\mathcal{F}}^2 + \|\mathbf{A}_{t_j}^{-1}\|_{\mathcal{F}}^2 \quad (6)$$

where  $\|\cdot\|_{\mathcal{F}}$  is the Frobenius norm. The accumulated discrete element energy functional of the soft body denotes

$$E(\mathbf{x}) = \sum_{t_j \in \mathcal{T}} \Psi(\mathbf{A}_{t_j}(\mathbf{x})) \quad (7)$$

where  $\mathbf{x} \in \mathbb{R}^{3 \times n}$  contains all discretized vertices location of the soft body  $\mathcal{M}$ .

3) *Rigidity-Aware AMH Constraints*: Monocular cameras are generally considered the primary sensory for environmental perception due to their ease of use and availability, compared to multiview systems. However, deformable shape reconstruction from 2-D image observations is well-known as an ill-posed inverse problem and has been actively researched [62]. We leverage the proposed volumetric discretized model and introduce rigidity-aware AMH constraints to make this problem trackable, aiming at reconstructing the soft finger's deformed shape reliably.

We model the mechanical coupling of the rigid plate for the fiducial marker in Fig. 1(a) as a uniform rigid transformation  $g$  for each attached node in the discrete model  $\mathcal{M}$ , as shown in Fig. 2(a) (iii) and (iv)

$$\mathbf{x}_h = g(\mathbf{X}_h) \quad (8)$$

where  $\mathbf{x}_h \in \mathbb{R}^{3 \times p}$  contains deformed location of  $p$  vertices related to the rigidity-aware AMH constraints while  $\mathbf{X}_h \in \mathbb{R}^{3 \times p}$  contains the corresponding undeformed vertices location. The rigid transformation  $g$  is estimated by fiducial markers widely used in robotic vision.

**Algorithm 1:** Projected Hessian Algorithm.

---

1: **Input:** Rigid transformation of AMH  $g$   
2: **Output:** Estimated positions of deformed vertices  $\mathbf{x}^*$   
**Require:**  
Vertices positions of current shape  $\mathbf{x}_0$   
Convergence tolerance  $\epsilon$   
Maximum number of iterations  $N_{\max}$

3:  $k \leftarrow 0$   
4: Compute gradient  $d_k = \nabla \tilde{E}(\mathbf{x}_k)$   
5: And Hessian  $H_k = \nabla^2 \tilde{E}(\mathbf{x}_k)$   
6: **while**  $\|d_k\| > \epsilon$  and  $k < N_{\max}$  **do**  
7:   Solve  $H_k \Delta \mathbf{x}_k = -d_k$  for  $\Delta \mathbf{x}_k$   
8:   Project  $\Delta \mathbf{x}_k$  onto the feasible region  
9:   Update iterate:  $\mathbf{x}_{k+1} \leftarrow \mathbf{x}_k + \Delta \mathbf{x}_k$   
10:    $k \leftarrow k + 1$   
11: **end while**  
12: iteration stop  $\mathbf{x}^* = \mathbf{x}_k$

---

4) *Geometric Optimization for Shape Estimation:* With the discrete energy function (7) of the given soft body  $\mathcal{M}$  and observed kinematics constraints (8), soft body shape estimation can be directly translated into a constrained geometry optimization problem as

$$\begin{aligned} \min_{\mathbf{x}} \quad & \sum_{t_j \in \mathcal{T}} \Psi(\mathbf{A}_{t_j}(\mathbf{x})) \\ \text{s.t.} \quad & \mathbf{x}_h = g(\mathbf{X}_h). \end{aligned} \quad (9)$$

Instead of considering kinematics constraints as hard boundary conditions, we enforce them by appending quadratic penalty terms to  $E(\mathbf{x})$  in (7) for easier handling, which results in

$$\tilde{E}(\mathbf{x}) = \sum_{t_j \in \mathcal{T}} \Psi(\mathbf{A}_{t_j}(\mathbf{x})) + \omega \|\mathbf{x}_h - g(\mathbf{X}_h)\|^2. \quad (10)$$

As illustrated in Fig. 2(a)(v), we can achieve deformed shape estimation by minimizing the augmented energy function in (10) as

$$\mathbf{x}^* = \arg \min_{\mathbf{x}} \tilde{E}(\mathbf{x}; \omega, g) \quad (11)$$

where  $\omega$  is the penalty parameter for the corresponding unconstrained minimization problem. A greater penalty weight will lead to better constraint satisfaction but poorer numerical conditions.

In practice, we set  $\omega = 10^5$  and compute the deformed vertices positions  $\mathcal{V}$  by iteratively minimizing (11) using a Newton-type solver shown in Algorithm 1. As shown in Fig. 2(b), a series of physically plausible deformations of the soft finger under observed constraints are reconstructed in real time using our proposed optimization approach.

### C. Object Shape Estimation Using Tactile Sensing

While proprioception refers to being aware of one's movement, tactile sensing involves gathering information about the external environment through the sense of touch. This section presents an object shape estimation approach by extending the

PropSE method proposed in the previous section to tactile sensing.

Since our soft finger can provide large-scale, adaptive deformation conforming to the object's geometric features through contact, we could infer shape-related contact information from the finger's estimated shape during the process. We assume the soft finger's contact patch coincides with that of the object during grasping. As a result, we can predict object surface topography using spatially distributed contact points on the touching interface.

1) *Contact Interface Points Extraction:* Based on the spatial discretization model in Section III-B1, an indexed set  $\mathcal{I} = \{c_1, c_2, \dots, c_k\}$  of nodes located at the upper area of the soft finger mesh  $\mathcal{M}$  are extracted as contact interface points, as shown in Fig. 3(a).

With each of the observed AMH constraints input, we could determine the positions of these contact interface points by first solving (11), then extracting corresponding nodes using indexed set  $\mathcal{I}$  by solving the deformed positions of vertices  $\mathcal{V}$ :  $\mathbf{x}_c = \{\mathbf{x}_i | \mathbf{x}_i \in \mathcal{V}, i \in \mathcal{I}\}$ .

2) *Implicit Surface Representation for Object Shape:* Considering the grasping action using a soft finger as a multipoint tactile probe, the object surface patches could be progressively reconstructed by these gripping actions with collected positions of contact interface points  $\mathbf{x}_c$  extracted from the soft finger.

An implicit surface representation is defined by a function that can be evaluated at any point in space, yielding a value indicating whether the point is inside the object, outside the object, or on the object's surface. For the 3-D space considered in our problem, this function  $f: \mathbb{R}^3 \rightarrow \mathbb{R}$  is defined as

$$f(\mathbf{x}) \begin{cases} < 0, & \text{if } \mathbf{x} \text{ inside the object} \\ = 0, & \text{if } \mathbf{x} \text{ on the surface} \\ > 0, & \text{if } \mathbf{x} \text{ outside the object.} \end{cases} \quad (12)$$

As is shown in Fig. 3(b), we only collected positions of partial contact interface points  $\mathbf{x}_c$ , which are assumed to coincide with the object surface for each gripping action. While surface points are observed, we do not explicitly observe off-surface or internal point exemplars. For those unobserved cases in (12), we generate control points of the corresponding two types to express the directional information of the surface using the method described in [63].

3) *GPIS for Surface Estimation:* An object's shape is estimated by finding the points with zero value of implicit surface function (12) (i.e., the isosurface) in the 3-D region of interest. The Gaussian process implicit surface (GPIS) method can be used as a tool for object surface reconstruction from partial or noisy 3-D data. It is a nonparametric probabilistic method often used for tactile and haptic exploration [64], [65].

A GP is a collection of  $N$  random variables with a joint Gaussian distribution which can be specified using its mean and covariance functions. The collected contact interface point and the generated control point positions  $\mathcal{X} = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N\}$  for each grasping action and the corresponding observed values are denoted as  $\mathcal{Y} = \{y_1, y_2, \dots, y_N\}$ . Here,  $y_i = f(\mathbf{x}_i) + \epsilon$ , where  $\epsilon \sim \mathcal{N}(0, \sigma_\epsilon^2)$  denotes Gaussian noise with zero mean

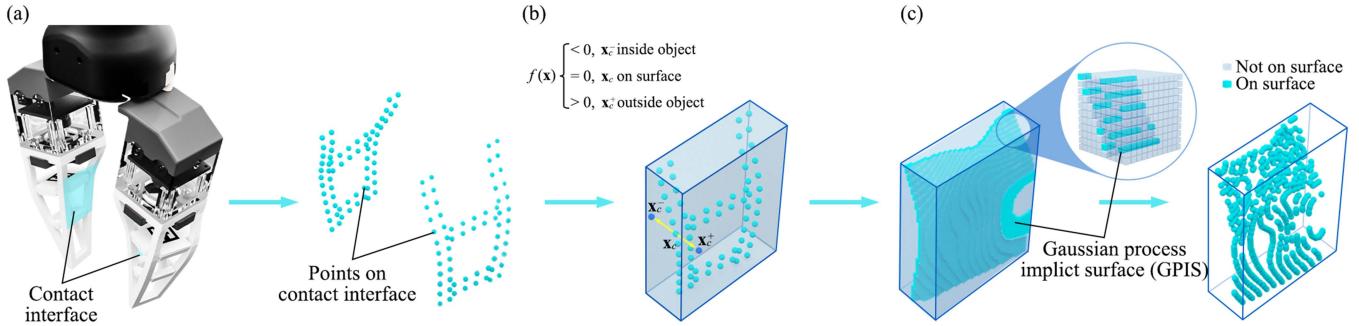


Fig. 3. Pipeline for contact interface geometry sensing using deformed positions of soft finger mesh nodes. (a) Contact interface points extraction: Because the soft finger can deform and adapt its shape to fit the contours of the object being grasped, we take the deformed soft finger mesh nodes as approximate multicontact points on the contact interface. (b) Implicit surface representation: In addition to the mesh nodes  $\mathbf{x}_c$  on the contact interface, auxiliary training points  $\mathbf{x}_c^-$  and  $\mathbf{x}_c^+$  are generated in this step to increase the accuracy of the implicit surface reconstruction. (c) GPIS for shape estimation: GPIS model is adopted for contact object surface patch estimation.

and  $\sigma_\epsilon^2$  variance. As a result, the GP can be written as  $f(\mathbf{x}) \sim \mathcal{GP}(m(\mathbf{x}), k(\mathbf{x}, \mathbf{x}'))$ , where  $m(\mathbf{x})$  is the mean function and  $k(\mathbf{x}, \mathbf{x}')$  is the covariance function [66].

In our implementation, we used the radial basis function kernel, which is characterized by the two hyperparameters, the variance  $\sigma_f^2$  and the length scale  $l$ , expressed as the following:

$$k(\mathbf{x}, \mathbf{x}') = \sigma_f^2 \exp\left(-\frac{\|\mathbf{x} - \mathbf{x}'\|^2}{2l^2}\right). \quad (13)$$

With the covariance function and the observation data, the predictive mean  $\bar{f}(\mathbf{x}^*)$  and variance  $\bar{V}(\mathbf{x}^*)$  at a query point  $\mathbf{x}^*$  are

$$\bar{f}(\mathbf{x}^*) = \mathbb{E}[f(\mathbf{x}^*) | \mathcal{X}, \mathcal{Y}, \mathbf{x}^*] = k(\mathcal{X}, \mathbf{x}^*)^T \Sigma \mathcal{Y} \quad (14)$$

$$\bar{V}(\mathbf{x}^*) = k(\mathbf{x}^*, \mathbf{x}^*) - k(\mathbf{x}^*, \mathcal{X})^T \Sigma k(\mathcal{X}, \mathbf{x}^*) \quad (15)$$

where  $\Sigma = (k(\mathcal{X}, \mathcal{X}) + \sigma_\epsilon^2 \mathcal{I})^{-1}$ . After voxelizing the bounding box volume enclosing the partially deformed finger-object interface, the zero-mean isosurface can be extracted from posterior estimation, which approximates the local shape of a grasped object, as is shown in Fig. 3(c).

## IV. RESULTS

### A. On Vision-Based PropSE

Here, we first present the benchmarking results against two widely adopted methods to demonstrate the superior performance of our proposed vision-based PropSE method. Then, we present the results of our proposed vision-based PropSE method using two experiment setups. One leverages motion capture markers as ground truth, providing high-precision but sparse measurements. The other uses a touch-haptic device for ground truth data collection, which is less accurate but contains larger measuring coverage on the soft finger.

The implementation of the proposed geometric optimization-based algorithm (Algorithm 1) was developed in C++ and evaluated on a computer with an Intel Core i7 3.8 GHz CPU and 16 GB of RAM. By leveraging the capabilities of

TABLE I  
RUN TIME AND MEAN ERROR COMPARISONS OF ABAQUS, ARAP, AND OUR METHOD

Number of Elements	1k	1.5k	3k	6k	12k	
Run Time (s)	Abaqus	16.7	17.1	18.7	22.1	27.8
	ARAP	0.0485	0.0776	0.151	0.304	0.703
	Ours	0.0231	0.0452	0.0905	0.241	0.639
Mean Error* (mm)	Abaqus	NA	NA	NA	NA	NA
	ARAP	0.857	0.902	0.754	0.878	0.926
	Ours	0.346	0.149	0.130	0.143	0.086

\*: The benchmark of mean error is Abaqus.

algorithmic differentiation within the numerical solver, *Eigen* [67], this system demonstrated the ability to compute deformations involving 1500 tetrahedra in real-time, achieving frame rates up to 20 FPS.

1) *Comparison With the Conventional Methods*: We performed a comparative analysis with two widely adopted techniques to showcase the efficacy of our shape estimation method. One is Abaqus, a premier finite element analysis (FEA) software extensively applied in structural analysis and deformation modeling across various engineering disciplines. This comparison aims to highlight the versatility and precision of our approach within contexts requiring intricate modeling capabilities. (Please refer to Appendix A for further details concerning the Abaqus simulation.)

The other is the ARAP method [68], a widely adopted method in digital geometry processing for estimating object shapes through minimal rigid deformation. This comparison is particularly valuable, as ARAP's principles of shape preservation align closely with the core objectives of our shape estimation task, providing valuable benchmarking. (Please refer to Appendix B for further details regarding our implementation.)

Table I compares our proposed method's run time and mean error with those mentioned earlier. Each method is evaluated on five meshes with increasing resolutions, resulting in 1, 1.5, 3, 6, and 12 k elements. The soft finger underwent six motions applied to the AMH shown in Fig. 2(b) with all the deformation data recorded. We treat the results from Abaqus as the ground truth. Results show that our method is 40 to 700 times faster



than Abaqus and 1 to 2 times faster than ARAP at different resolutions. We also compared the mean errors of all nodes estimated by our method and ARAP when benchmarked against Abaqus. The results show that our method's mean error decreases significantly, from 0.346 to 0.086 mm, as the number of elements increases. The ARAP's error ranges from 0.7 to 1.0 mm for different meshes. Our approach shows significant advantages over Abaqus and ARAP regarding running time and accuracy.

The optimization solver deployed to minimize the ARAP energy leverages the local/global method (as detailed in Appendix B). While this solver efficiently approximates the local minimum, its approach to convergence toward a numerical minimum necessitates a considerable number of iterations, a characteristic underscored during implementation [61]. We fixed the number of iterations at 10 for our benchmarking procedure to achieve convergence. This predefined iteration limit could account for the observed comparative slowness of the ARAP optimization solver relative to our proposed method. Regarding the evaluation of mean error, the suboptimal performance of ARAP, as compared to ours, might be attributed to the local/global optimization solver settings. Moreover, the deformation energy model used by ARAP might not fully encompass the nonlinear deformation behaviors of our soft robotic fingers.

We also observe that the error of our method decreases most dramatically when the number of elements increases from 1 to 1.5 k, and the error reduction from 1.5 to 6 k is marginal. Hence, the mesh with 1.5 k elements is the most appropriate for our method, achieving both faster run speed and minor error, which was selected for real-time estimation in the following experiments. (Please refer to Appendix C for additional results on Algorithm 1 parameter.)

2) *Deformation Estimation With Motion Capture Markers*: Shown in Fig. 4(a) is the soft robotic finger mounted on a three-axis motion platform for interactive deformation estimation. The test platform is operated manually to generate a set of contact configurations between the soft finger and the indenter. During the process, the in-finger camera streams real-time image data at a resolution of  $640 \times 480$  pixels. Using an off-the-shelf ArUco detection library, the detected AMH rigid motion is fed into our implemented program for deformation estimation.

A motion capture system (Mars2H by Nokov, Inc.) was used to track finger deformations through nine markers with an 8 mm radius. Among them, six markers were divided into three pairs, which were rigidly attached to the fingertip ( $m_5, m_6$ ), the first layer ( $m_3, m_4$ ), and the second layer ( $m_1, m_2$ ) of the soft finger, respectively. The other three markers were attached to the platform and used as the reference reading to align the motion capture system's reference frame with the platform's coordinate frame.

The markers were attached to the soft finger with rigid links, as shown in Fig. 4(b). We designed the connecting links to be three lengths to avoid occlusion during tracking. We assume each marker is rigidly attached to the nearest tetrahedron on the parameterized mesh model  $\mathcal{M}$ , representing the estimated marker location using barycentric coordinates of the corresponding tetrahedron element in the soft robotic finger's deformed

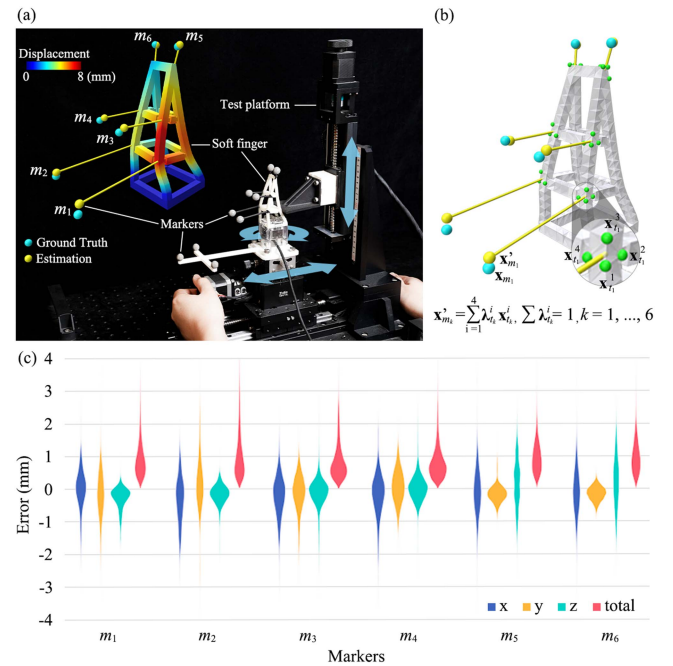


Fig. 4. Estimated marker deformation obtained by the proposed PropSE method. (a) Experimental setup, including the soft finger, embedded in an RGB camera, a manual three-axis motion test platform, and six motion capture markers  $m_1, m_2, \dots, m_6$ , rigidly attached to the soft finger. (b) Estimated position of the marker  $x'_{m_k}$  is calculated using the barycentric coordinate of the corresponding attached tetrahedron  $t_k$ , while the ground truth reading  $x_{m_k}$  is obtained from the motion capture system. (c) Corresponding error for each marker's 3-D deformation and total norm.

states

$$\mathbf{x}'_{m_k} = \sum_{i=1}^4 \lambda_{t_k}^i \cdot \mathbf{x}_{t_k}^i, k \in \{1, 2, \dots, 6\} \quad (16)$$

$$\sum_{i=1}^4 \lambda_{t_k}^i = 1, t_k \in \mathcal{T}. \quad (17)$$

Due to the rigid connection assumption, the barycentric coordinates  $\lambda_{t_k}$  are constant during deformation. We solve the barycentric coordinates in (16) using the tetrahedron's initial vertex position and corresponding tracked marker position without contact. The marker position prediction model is a linear combination of the deformed vertex position of the corresponding tetrahedron resulting from geometric optimization in Algorithm 1 using calibrated barycentric coefficients. (See Movie S1 in the Supplementary Materials for a video demonstration.)

We visualize the error distribution with 3 k pairs of the six markers' estimated and ground truth positions as illustrated in Fig. 4(c). The norm of the six markers' total error is within 3 mm, while the error distribution along each axis is centered around the  $(-2, 2)$  mm range. As the marker prediction model in (16) comprises calibration and geometric optimization, the error distribution of six sparse markers may only partially validate the proposed method, leading to the next experiment.

3) *Deformation Estimation Using Touch Haptic Device*: We designed another validation experiment using the pen-nib's

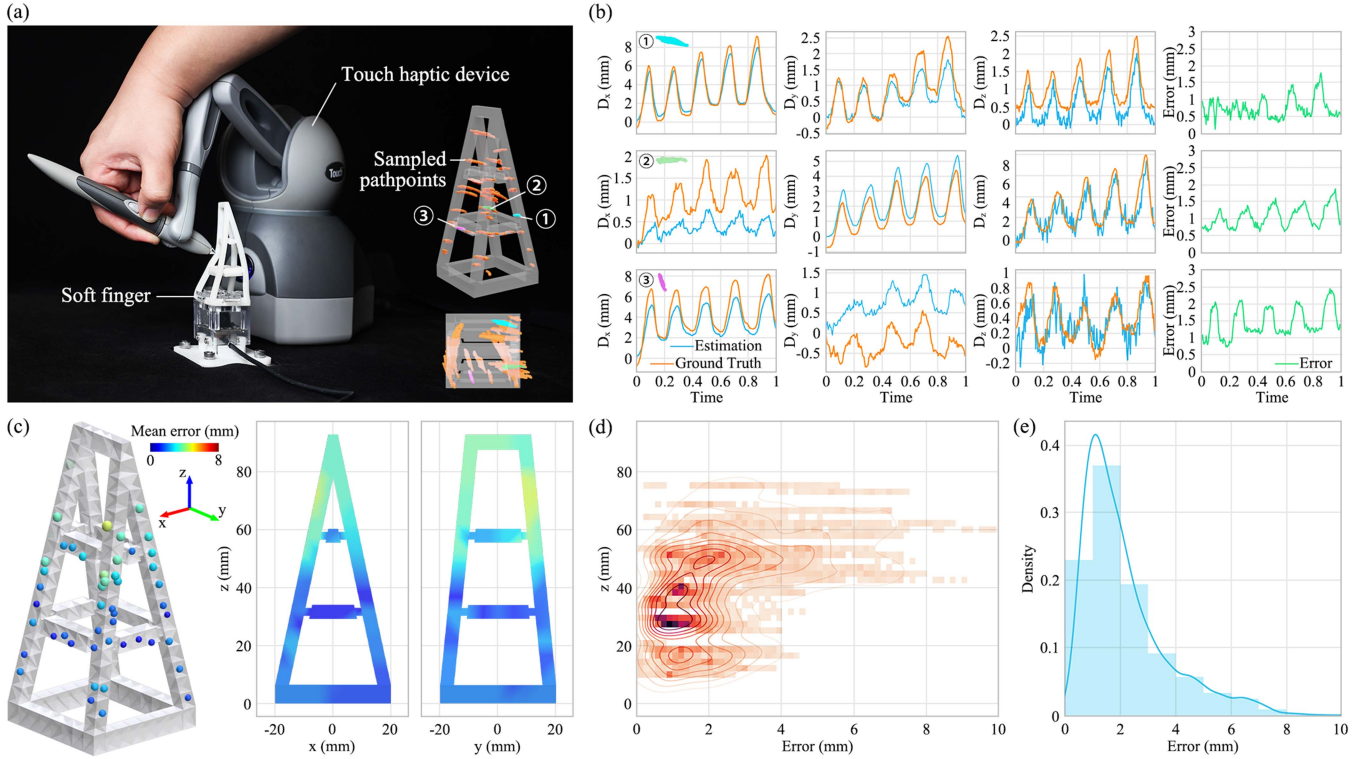


Fig. 5. Estimated deformation field of the soft finger using the PropSE method. (a) Touch haptic device is used to make contact with the soft finger at different locations while simultaneously recording the ground-truth positions and the reconstructed positions of contact points. (b) Three sampled pushing trajectories of the pen-nib and corresponding measurements from the PropSE method. Total errors are reported in the last column. The pen-nib of the touch haptic device is pushed forward and backward five times at each location. (c) Fifty testing locations sampled are spread over half of the side of the soft finger. The mean error norm map is interpolated using the values of the fifty sampled contact locations. (d) Distribution of the total errors along the height ( $Z$ -axis) of the soft finger. (e) Distribution of the total errors of sampled contact points.

position of a haptic device (Touch by 3D Systems, Inc.) as ground truth measurement. As shown in Fig. 5(a), an operator holding the pen-nib initiated contact at a random point on the soft robotic finger by pushing it five times. Fifty points were sampled, spreading over half of the soft robotic finger with recorded pen-nib position and the corresponding point of contact on the estimated deformation in the mesh model.

Similar to the calibration process when using the motion capture system, we solve the barycentric coordinates in (16) using the initial contact position of pen-nib and the undeformed vertex position of the tetrahedron nearest to the contact point. Since there is no slipping between the contact point and the pen-nib, recording the pushing position of the pen-nib for a randomly selected point is equivalent to collecting the ground truth deformation field of the soft finger evaluated at that point. Fig. 5(b) shows three selected pushing trajectories and the corresponding errors between estimation and ground truth. The pushing duration lasts around ten seconds for each location and is rescaled to 1 in the plot. The data is recorded at 20 Hz. Due to the variations among the pushing trajectories among the three locations, the errors are slightly different, but all lie within a 2.5 mm range.

The haptic device measurements cover an extensive portion of the soft robotic finger, revealing further details regarding the spatial distribution of the estimation errors. We visualize the mean errors of deformation estimation evaluated at

the fifty randomly selected contact locations in Fig. 5(c). We interpolated two side views of continuous error distribution for the soft robotic finger with errors of all sampled locations using a Gaussian-kernel-based nearest-neighbor method [69].

Contact locations near the observed AMH constraint are expected to exhibit fewer errors due to penalized computation near this region during deformation optimization. We plot the error distribution of all sampled locations along the  $Z$ -axis in Fig. 5(d). Contact locations with a similar height to the AMH constraint exhibit a smaller and more concentrated error distribution. Fig. 5(e) shows the error histogram of the overall experiment records, where the median of estimated error for the whole-body deformation is 1.96 mm, corresponding to 2.1% of the finger's length. (See Movie S2 in the Supplementary Materials for a video demonstration.)

### B. On Amphibious Tactile Sensing for PropSE

Here, we further investigate our proposed method in amphibious tactile sensing through three experiments in lab conditions. We begin by benchmarking our proposed VBTS method at controlled turbidity underwater. Then, we present a touch-based object shape reconstruction task to demonstrate the application of our proposed solution for amphibious tactile sensing. Finally, we present a full-system demonstration by attaching our robotic finger to the gripper of an underwater remotely operated vehicle



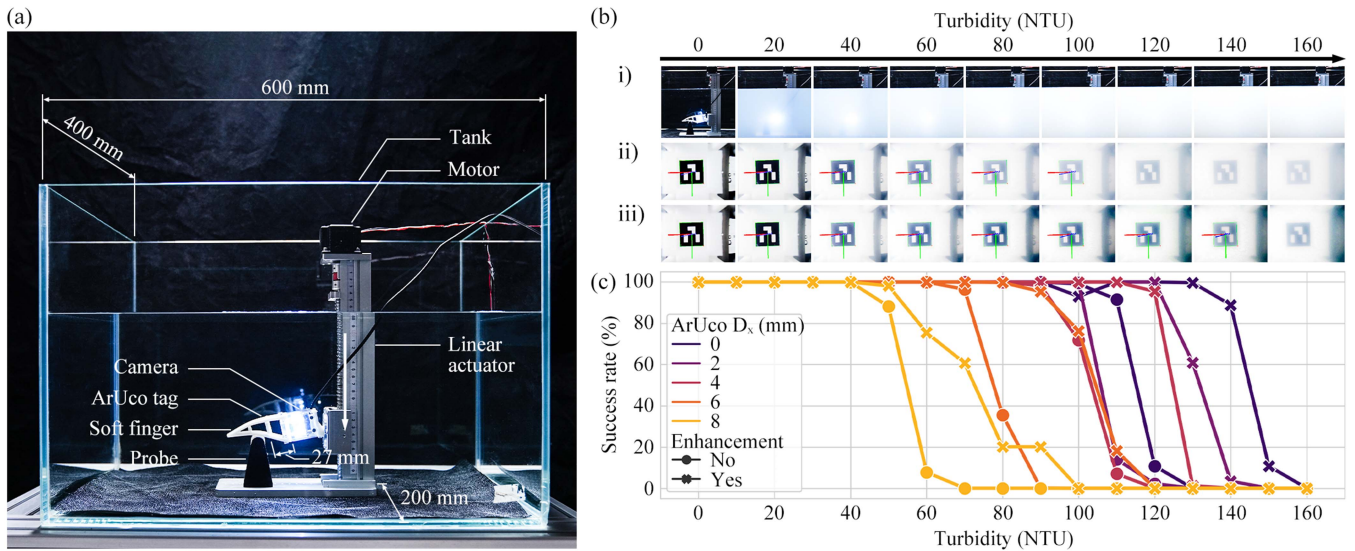


Fig. 6. Benchmarking results in different turbidity conditions underwater in a lab tank. (a) Experiment was set up in a room with controlled ambient lighting of 3000 lumens placed atop the tank (not shown in this picture). (b) Images taken by adding condensed standard turbidity liquid to increase the water turbidity from 0 to 160 NTU, including i) experiment pictures taken by an external camera at the same angle as (A); ii) raw images captured by the in-finger vision overlaid with triad coordinates to indicate successful pose recognition; and iii) digitally enhanced images overlaid with triad coordinates to indicate successful pose recognition. (c) Results on the pose recognition success rate of the ArUco marker from the in-finger vision under increasing tank turbidity when pushing the soft robotic finger at different target displacements, with or without image enhancement.

(ROV) for underwater grasping in a water tank, which we plan to implement further in the field test soon.

1) *Benchmarking VBTS Underwater Against Turbidity*: Our proposed rigidity-aware AMH method effectively transforms the visual perception process for deformable shape reconstruction into a marker-based pose recognition problem. Therefore, the benchmarking of our VBTS solution underwater is directly determined by successfully recognizing the fiducial marker poses used in our system under different turbidity conditions. Turbidity is an optical characteristic that measures the clarity of a water body and is reported in NTU [70]. It influences the visibility of optical cameras for underwater inspection, inducing light attenuation effects caused by the suspended particles [71]. As one of the critical indicators for characterizing water quality, there have been rich studies on the turbidity of large water bodies worldwide. For example, Li et al. [72] showed that the Yangtze River's turbidity was measured between 1.71 and 154 NTU.

We investigated the robustness of our proposed VBTS solution in different water clarity conditions by mixing condensed standard turbidity liquid with clear water to reach different turbidity ratings. Fig. 6(a) shows the experiment setup. Our proposed soft robotic finger is installed on a linear actuator in a tank filled with 56 liters of clear water. A probe is fixed under the soft robotic finger, inducing contact-based whole-body deformation when the finger is commanded to move downward. The tank is placed in a room with controlled ambient lighting of 3000 lumens placed atop the tank. We controlled the linear actuator for each turbidity condition so that the finger moved downward along the  $x$ -axis. This enabled us to record the ArUco image streams when fixed 0, 2, 4, 6, and 8 mm displacements in  $D_x$  are reached. For example, the three images shown in the first column of Fig. 6(b) are i) the experiment scenario taken at the same angle as Fig. 6(a) when the turbidity is zero (before

adding condensed standard turbidity liquid), ii) a sample of the raw image captured by our soft robotic finger's in-finger camera, and iii) image enhancement based on the image shown in ii), respectively. The water tank's clarity is modified by adding specific portions of condensed standard turbidity liquid to reach different turbidity ratings at 10 NTU per step (images for 20 NTU per step increase are shown in Fig. 6(b) for the ease of visualization), increasing from 0 to 160 NTU, covering the Yangtze River's turbidity range.

For each of the  $D_x$  positions, we recorded 1000 images using our soft robotic finger's in-finger camera to obtain the pose recognition success rate (%) under each turbidity rating, before or after image enhancement, reported in Fig. 6(c). The results reported in Fig. 6(c) involve data aggregated from 85 000 raw images (1000 images per NTU step per ArUco position  $\times$  17 NTU steps  $\times$  5 ArUco positions) from in-finger vision for ArUco pose recognition, which is doubled after image enhancement, resulting a total of 170 K images.

In our experiment, for the turbidity range between 0 and 40 NTU, the raw images captured by our in-finger vision achieved a 100% success rate in ArUco pose recognition. At 50 NTU turbidity, the first sign of failed marker pose recognition was observed when the most considerable deformation was induced at 8 mm of  $D_x$ . Our experiment shows that this issue can be alleviated using simple image enhancement techniques to regain a 100% marker pose recognition success rate. However, the marker pose recognition performance under large-scale whole-body deformation quickly deteriorated when the turbidity reached 60 NTU and eventually became unusable at 70 NTU. Image enhancement could effectively increase the upper bound to 100 NTU to reach an utterly unusable marker pose recognition in large-scale whole-body deformation. However, for small or medium whole-body deformations measured by  $D_x \leq 6$  mm,

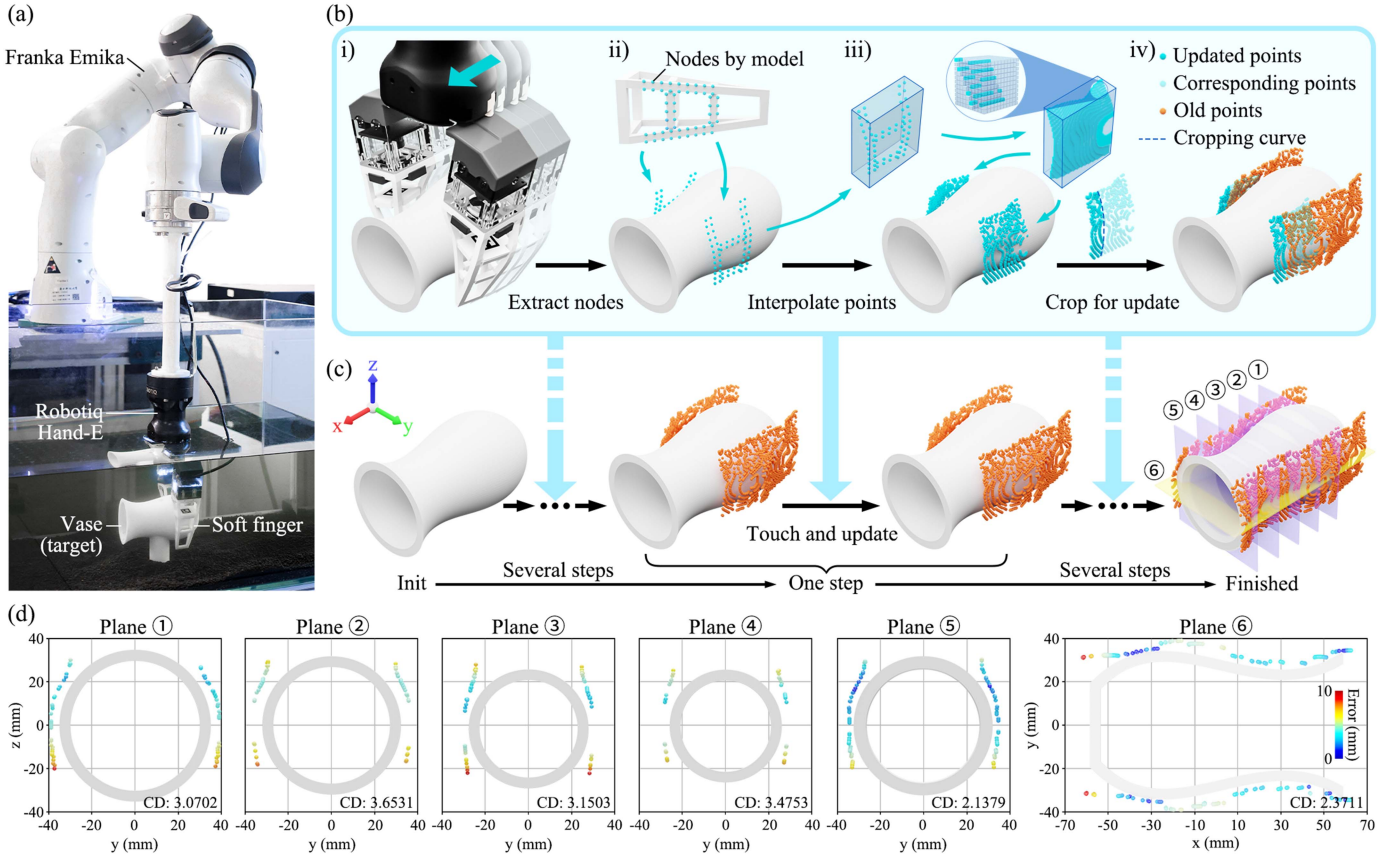


Fig. 7. Underwater shape estimation of a vase using PropSE of the soft finger. (a) Experimental setup for underwater shape estimation. A Robotiq Hand-E gripper, installed with two proprioceptive soft fingers and an extension link, is mounted on a Franka Emika Panda robot arm. The gripper is programmed to periodically perform a series of actions, including gripping, releasing, and moving along the x-axis for a fixed distance. At the same time, a vase is fixed at the bottom of the tank in the lab. (b) Contact surface patch prediction using GPIS with the soft finger. (c) Experiment pipeline for underwater shape estimation of a vase. (d) Evaluation of the reconstructed vase shape on some cutting sectional planes, measured in CD.

our system remains functional until around 100 NTU in turbidity, where simple image enhancement techniques help for a balanced consideration between algorithmic cost, engineering complexity, and system performance.

For turbidity above 100 NTU, simple image enhancement provides limited contributions to our system. Our experiment shows that when the turbidity reached 160 NTU, our in-finger system failed to recognize any ArUco pose underwater, even after image enhancement. Since blurry images of the marker remain visible in the captured images, we can: 1) use more advanced image processing algorithms; 2) use better imaging hardware; 3) apply stronger ambient lighting; and 4) redesign the marker pattern specifically for underwater usage to systematically increase the upper bound of the turbidity rating for marker-based pose estimation in contact-based amphibious grasping using VBTS methods. Results obtained from this experiment provide a general understanding of the potential regions for amphibious grasping characterized by turbidity with possible solutions to improve further.

2) *Underwater Exteroceptive Estimation of Object Shape:* In this experiment, we apply our soft robotic finger with in-finger vision to a contact-based shape reconstruction task to demonstrate our solution's capabilities in underwater exteroceptive estimation. Shown in Fig. 7(a) is the experimental setup

conducted in the lab condition using the same water tank as the one used in the previous experiment. In this case, we used a parallel two-finger gripper (HandE from Robotiq, Inc.) attached to the wrist flange of a robotic manipulator (Franka Emika) through a 3-D-printed cylindrical rod for an extended range of motion. Our soft robotic fingers are attached to each fingertip of the gripper through a customized adapter fabricated by 3-D printing. Our previous work extensively tested this IP67 gripper's underwater servoing capabilities for reactive grasping during temporary submergence under the water [73]. In this study, we use the same gripper for underwater object shape estimation in a lab tank. One can always replace the Hand-E gripper with a professional underwater gripper for more intensive underwater usage in the field.

With the gripper submerged underwater, the system is programmed to sequentially execute a series of actions, including gripping and releasing the object and moving along a prescribed direction for a fixed distance to acquire underwater object shape information, as shown in Fig. 7(b)(i). By mounting the target object at the bottom of the tank, we assume that: 1) the object's pose is fixed and calibrated with the gripper and 2) passive object shape exploration is considered for object coverage. The inference of the GPIS model is computationally intractable for the large  $N$  measurement that accrues from high-dimensional

tactile measurements. Instead of predicting the whole object surface by accumulating all the collected data, we only query a local GPIS model approximated using current observed contact data in a local focus area and build the surface incrementally, as shown in Fig. 7(b)(ii) and (iii).

*a) Local GPIS Model Inference:* A training set containing contact interface points  $\mathbf{x}_c$  and corresponding augmented control points are collected each time a grasping action is performed. Before querying the local GPIS model in the interested area, hyperparameters  $\sigma_f^2$  and  $l$  associated to (13) are optimized first using the standard training method for GPs, i.e., maximizing the marginal likelihood. Then, we evaluate the local GP on voxel grid points at a resolution of 0.2 mm in the interested area and keep those points with zero mean of (14) as estimated points on the surface patch of the object.

*b) Local Patches Concatenation:* After calibrating the object pose to the gripper, we programmed the grasping system to follow a predefined path for object shape exploration. As is shown in Fig. 7(b)(iv), each time after GPIS query in the local 3-D region, a global registration action is performed by transforming these local iso-surface points into the global space. Leveraging the continuous nature of the predefined exploration path, a simple surface concatenation strategy is used, where only the points of the estimated surface patch corresponding to moving distance are kept, and points of overlapping intervals belonging to the latest estimated surface patch are rejected. As is shown in Fig. 7(c), after initialization of the relative pose between the gripper and the object, the shape of the object is continuously reconstructed using the described passive exploration strategy.

*c) Object Shape Estimation Evaluation:* In Fig. 7(d), we present our method on actual data collected during the underwater tactile exploration experiment. The shape estimates at each cutting sectional plane are compared concerning the ground truth using the Chamfer distance (CD) [74], a commonly used shape similarity metric. We chose five vertical cutting planes and one horizontal sectional plane for reconstructed object surface evaluation. For each cutting plane, a calibration error exists between the vase and the Hand-E gripper, leading to the expected gap between the reconstructed and ground truth points. In addition to the systematic error, we have observed a slight decrease in the CD metric values between planes 1 and 5 compared to planes 2, 3, and 4, which could be attributed to the limitations of the soft finger in adapting to small objects with significant curvature. On the other hand, by employing tactile exploration actions with a relatively large contact area on the soft finger's surface, the shape estimation of objects similar in size to the vase can be accomplished more efficiently, typically within 8–12 touches. The 3-D-printed vase has dimensions of approximately 80 mm by 80 mm by 140 mm. (See Movie S3 in the Supplementary Materials for a video demonstration.)

*3) Vision-Based Tactile Grasping With an Underwater ROV:* Here, we provide a full-system demonstration using our vision-based soft robotic fingers on an underwater remotely operated vehicle. (ROV, FIFISH PRO V6 PLUS by QYSEA.)<sup>2</sup> It includes

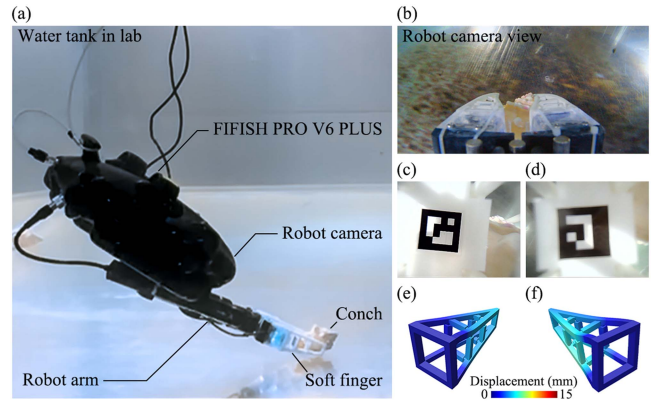


Fig. 8. Demonstration of our soft robotic finger with in-finger vision for tactile sensing underwater. (a) Key components involved in the test. (b) Screenshot of a 4K image captured by the underwater ROV's onboard camera when our fingers are holding a conch after successfully grasping. (c) and (d) Screenshot of the images captured by an in-finger vision camera in the left and right fingers while holding the conch. (e) and (f) Whole-body deformation reconstruction for both fingers based on the images captured by the in-finger vision cameras, respectively.

a single-DOF robotic gripper, which can be modified using the proposed soft fingers with customized adaptors.

The experiment results reported in Section IV-B1 already benchmark our system's promising capabilities for real-time underwater tactile sensing. As shown in Fig. 6(b), the water at 20 NTU or above is already very challenging to observe from a third-personal perspective. Experiments underwater would require additional cost to prepare a second underwater ROV to record videos when the water is clear enough. However, as analyzed above, our in-finger vision could perform nicely at a much higher NTU range. Therefore, in this section, we only conducted this experiment in a lab tank to demonstrate our system's integration with an existing underwater ROV system during an underwater task.

Shown in Fig. 8(a) is a brief overview of the system and the scene. Our fingers are attached to the underwater ROV's gripper through 3-D-printed adaptors to replace the default rigid fingers. Our design conveniently introduced omnidirectional adaptation capability to the gripper's existing functionality with added capabilities in real-time tactile sensing underwater. Shown in Fig. 8(b) is a screenshot of the image taken by the ROV's onboard camera, recording 4K videos in real-time. In this experiment, both soft robotic fingers are installed with in-finger vision, capturing images shown in Fig. 8(c) and (d). Using these in-finger images, we can use the methods proposed in this work to achieve real-time reconstruction of contact events on our soft robotic finger in Fig. 8(e) and (f), while performing grasping tasks underwater.

See Movie S4 in the Supplementary Materials for a video demonstration. Besides the capabilities demonstrated in this article, we also identified an interesting observation during the experiment, adding to the benefits of having soft robotic fingers for underwater ROVs compared to the traditional rigid ones. While performing grasping underwater, the target objects are usually at the bottom. It is challenging for the underwater ROV to approach the target object smoothly and slowly, even in the lab tank with no water disturbances, which is also highly

<sup>2</sup><https://www.qysea.com/>



related to the pilot skills. Our soft fingers offer an added layer of production when the fingers collide with the bottom or other obstacles underwater, providing impact absorption for the underwater ROV while providing capable grasping and tactile sensing capabilities. If the original rigid fingers were installed, sudden impacts would occur when a collision happens, causing damage to the robot, the finger and gripper, and the underwater environment.

## V. DISCUSSION

### A. Encoding Large-Scale, Whole-Body Deformation by Tracking a Single Visual Representation

This study presents a model-based representation by tracking a single visual feature to achieve high-performing reconstruction of large-scale, whole-body deformation for PropSE. We introduced rigidity-aware AMH constraints during the modeling process. This problem is usually characterized by infinite degrees of freedom (DOFs) via a single visual feature in a 6-D pose. As a result, we effectively reduced the dimensionality in representing soft, large-scale, whole-body deformation. Our method shows 40 to 700 times faster run-time than commercial software such as Abaqus at different resolutions while exhibiting superior accuracy in deformation reconstruction. Our method also shows 1 to 2 times faster than the widely adopted ARAP algorithm. It should be noted that it remains theoretically unsolved to provide a model-based explicit proof regarding this problem, requiring further research in future works. However, our study shows promising capabilities of this approach toward a high-performing solution with real-time reconstruction efficiency and accuracy that can be used for tactile robotics.

### B. Rigid-Soft Interactive Representation in Tactile Robotics

The guiding principle behind our solution is a physical representation process shared by many existing solutions in VBTS technologies. Robotics usually interprets the physical world as an object-centric environment, which can be modeled as rigid bodies, soft bodies, or realistic bodies depending on predefined assumptions. A critical task in robotics is to provide a structured, digitalized representation of the unstructured, physical interactions so that the robotic system can make reliable action plans. The various designs of the soft medium in VBTS generally function as a *physical filter* to transform unstructured, object-centric properties from the external environment into a constrained problem space within the finger toward a refined representation. In this study, we propose a rigid-soft interactive representation using a rigid body (the marker plate) attached to the soft body (the adaptive finger) during contact-based interactions (filled with realistic bodies with various material stiffness). This process is similar to the mass-point model in physics, which provides a succinate placeholder for deriving various physical properties without losing generality in the mathematical formulation. Further development following such representation principle may give researchers a novel perspective to model

robotic dynamics as a tactile network of rigid-soft interactive representations, as demonstrated by results reported in this study.

### C. Vision-Based Multimodal Tactile Sensing for Robotics

In this study, we focus our investigation in VBTS on deformation reconstruction only, which can further implement tactile sensing of other perceptual modalities, as demonstrated in our previous work. For example, our recent work [14] achieved state-of-the-art performance in 6-D force-and-torque (FT) estimation using a similar design, where a fiducial marker is also attached inside the finger to provide a convenient representation. Combining both methods will achieve a vision-based multimodal tactile Sensing system in our soft robotic finger design, simultaneously providing high-performing tactile sensing in 6-D FT and continuous whole-body deformation reconstruction. This will address a significant challenge in robot learning from demonstration [75], [76], [77]. Wan and Song [78] also showed the possibility of achieving object detection in the external environment using the in-finger vision with a markerless design by implementing the in-painting technique. Our research provides a comprehensive demonstration regarding the robotic potentials of VBTS technology in fundamental theory and engineering applications, contributing to tactile robotics as a promising direction for future research [25].

### D. VBTS for Amphibious Robotics

Another novelty of this study is the application of VBTS in amphibious robotics. Our study presents comprehensive results and demonstrations in benchmarking performances, shape reconstruction tasks, and system integration with an underwater remotely operated vehicle. Many VBTS solutions require a closed chamber for the miniature camera to implement the photometric principle for tactile sensing, which may become challenging or even unrealistic for a direct application underwater. It should be noted that even after filling the closed chamber with a highly transparent resin to seal the camera, the layer of soft material used on the contact surface needs a depth-dependent calibration that is unrealistic to perform underwater. Furthermore, the soft material, such as silicon gel, will become brittle as the water depth increases [79]. Our previous work already showcased the engineering benefits of our soft robotic finger design, which can be used to reliably estimate 6-D FT from on-land to underwater scenarios [73]. In this work, we further demonstrate the applications of VBTS in high-performing shape reconstruction through our soft robotic finger design for amphibious applications. Our soft finger's metamaterial network leverages structural adaptation by design instead of being solely dependent on the material softness. This significantly reduces the fluidic pressure on our finger's adaptive behavior. Further discussion of this topic is outside the scope of this study, which we will address in an upcoming work with more details.

## VI. CONCLUSION

In conclusion, this study presented a novel VBTS approach for proprioceptive state estimation focusing on amphibious applications. Utilizing an SPN structure coupled with marker-based in-finger vision, our method achieved real-time, high-fidelity tactile sensing that accommodated omnidirectional adaptations. Introducing a model-based approach with rigidity-aware AMH constraints enabled effective optimization of the soft robotic finger's deformation. Furthermore, restructuring our proposed approach as an implicit surface model demonstrated superior shape reconstruction and touch-point estimation performance compared to existing solutions. Experimental validations affirmed its efficacy in large-scale reconstruction, turbidity benchmarking, and tactile grasping on an underwater ROV, thereby highlighting the potential of tactile robotics for advanced amphibious applications.

However, the study had several limitations. Manufacturing inconsistencies inherent to soft robots can impact the accuracy of our method, and algorithmic parameters required precise calibration through physical experiments. Additionally, using a rigid plate for boundary condition acquisition slightly hampered the finger's compliance, affecting the contact-based conformation between the object and our finger. The object surface estimation pipeline was also sensitive to contact geometry, restricting its use to local surface patches with smooth curvature changes.

Future research aims to optimize the system for versatile tactile grasping and expand its integration into robotic grippers for diverse on-land and underwater applications. The vision-based proprioception method holds the potential for developing advanced robotic necks for underwater humanoids with precise state estimation driven by parallel mechanisms or pneumatic actuation. These advancements will pave the way for the broader application and utility of VBTS technologies in robotic systems operating in complex environments.

## REFERENCES

- [1] A. Billard and D. Kragic, "Trends and challenges in robot manipulation," *Science*, vol. 364, no. 6446, 2019, Art. no. eaat8414.
- [2] F. Díaz Ledezma and S. Haddadin, "Machine learning-driven self-discovery of the robot body morphology," *Sci. Robot.*, vol. 8, no. 85, 2023, Art. no. eadh0972.
- [3] I. Van Meerbeek, C. De Sa, and R. Shepherd, "Soft optoelectronic sensory foams with proprioception," *Sci. Robot.*, vol. 3, no. 24, 2018, Art. no. eaau2489.
- [4] R. S. Dahiya, G. Metta, M. Valle, and G. Sandini, "Tactile sensing—From humans to humanoids," *IEEE Trans. Robot.*, vol. 26, no. 1, pp. 1–20, Feb. 2010.
- [5] R. S. Johansson and J. R. Flanagan, "Coding and use of tactile signals from the fingertips in object manipulation tasks," *Nature Rev. Neurosci.*, vol. 10, no. 5, pp. 345–359, 2009.
- [6] S. Sundaram, P. Kellnhöfer, Y. Li, J.-Y. Zhu, A. Torralba, and W. Matusik, "Learning the signatures of the human grasp using a scalable tactile glove," *Nature*, vol. 569, no. 7758, pp. 698–702, 2019.
- [7] S. Chun et al., "An artificial neural tactile sensing system," *Nature Electron.*, vol. 4, no. 6, pp. 429–438, 2021.
- [8] B. Shih et al., "Electronic skins and machine learning for intelligent soft robots," *Sci. Robot.*, vol. 5, no. 41, 2020, Art. no. eaaz9239.
- [9] R. S. Dahiya, P. Mittendorfer, M. Valle, G. Cheng, and V. J. Lumelsky, "Directions toward effective utilization of tactile skin: A review," *IEEE Sensors J.*, vol. 13, no. 11, pp. 4121–4138, Nov. 2013.
- [10] O. M. Andrychowicz et al., "Learning dexterous in-hand manipulation," *Int. J. Robot. Res.*, vol. 39, no. 1, pp. 3–20, 2020.
- [11] Z. Kappassov, J.-A. Corrales, and V. Perdereau, "Tactile sensing in dexterous robot hands," *Robot. Auton. Syst.*, vol. 74, pp. 195–220, 2015.
- [12] H. Yousef, M. Boukallel, and K. Althoefer, "Tactile Sensing for Dexterous In-Hand Manipulation in Robotics — A Review," *Sensors Actuators A: Phys.*, vol. 167, no. 2, pp. 171–187, 2011.
- [13] F. Wan, X. Liu, N. Guo, X. Han, F. Tian, and C. Song, "Visual learning towards soft robot force control using a 3D metamaterial with differential stiffness," in *Proc. Conf. Robot Learn.*, 2022, pp. 1269–1278.
- [14] X. Liu, X. Han, W. Hong, F. Wan, and C. Song, "Proprioceptive learning with soft polyhedral networks," *Int. J. Robot. Res.*, pp. 1–20, 2024, doi: [10.1177/02783649241238765](https://doi.org/10.1177/02783649241238765).
- [15] T. Kim, S. Lee, T. Hong, G. Shin, T. Kim, and Y.-L. Park, "Heterogeneous sensing in a multifunctional soft sensor for human-robot interfaces," *Sci. Robot.*, vol. 5, no. 49, 2020, Art. no. eabc6878.
- [16] F. Faure et al., *SOFA: A Multi-Model Framework for Interactive Physical Simulation*. Berlin, Germany: Springer, 2012, pp. 283–321.
- [17] H. Stuart, S. Wang, O. Khatib, and M. R. Cutkosky, "The ocean one hands: An adaptive design for robust marine manipulation," *Int. J. Robot. Res.*, vol. 36, no. 2, pp. 150–166, 2017.
- [18] S. Suresh, M. Bauza, K.-T. Yu, J. G. Mangelson, A. Rodriguez, and M. Kaess, "Tactile SLAM: Real-time inference of shape and pose from planar pushing," in *Proc. IEEE Int. Conf. Robot. Autom.*, 2021, pp. 11322–11328.
- [19] A. Mazzeo et al., "Marine robotics for deep-sea specimen collection: A systematic review of underwater grippers," *Sensors*, vol. 22, no. 2, 2022, Art. no. 648.
- [20] Q. Li, O. Kroemer, Z. Su, F. F. Veiga, M. Kaboli, and H. J. Ritter, "A review of tactile information: Perception and action through touch," *IEEE Trans. Robot.*, vol. 36, no. 6, pp. 1619–1634, 2020.
- [21] G. De Maria, C. Natale, and S. Pirozzi, "Force/Tactile sensor for robotic applications," *Sensors Actuators A: Phys.*, vol. 175, pp. 60–72, 2012.
- [22] E. Magrini, F. Flacco, and A. De Luca, "Estimation of contact forces using a virtual force sensor," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, 2014, pp. 2126–2133.
- [23] R. Holladay, T. Lozano-Pérez, and A. Rodriguez, "Planning for multi-stage forceful manipulation," in *Proc. IEEE Int. Conf. Robot. Automat.*, 2021, pp. 6556–6562.
- [24] H.-C. Lin and M. Mistry, "Contact surface estimation via haptic perception," in *Proc. IEEE Int. Conf. Robot. Autom.*, 2020, pp. 5087–5093.
- [25] S. Haddadin, L. Johannsmeier, and F. D. Ledezma, "Tactile robots as a central embodiment of the tactile internet," *Proc. IEEE*, vol. 107, no. 2, pp. 471–487, Feb. 2019.
- [26] Y. Yan, Y. Shen, C. Song, and J. Pan, "Tactile super-resolution model for soft magnetic skin," *IEEE Robot. Autom. Lett.*, vol. 7, no. 2, pp. 2589–2596, Apr. 2022.
- [27] Y. Wu et al., "A skin-inspired tactile sensor for smart prosthetics," *Sci. Robot.*, vol. 3, no. 22, 2018, Art. no. eaat0429.
- [28] F. Liu et al., "Printed synaptic transistor-based electronic skin for robots to feel and learn," *Sci. Robot.*, vol. 7, no. 67, 2022, Art. no. eabl7286.
- [29] Y. Yan et al., "Soft magnetic skin for super-resolution tactile sensing with force self-decoupling," *Sci. Robot.*, vol. 6, no. 51, 2021, Art. no. eabc8801.
- [30] W. Yuan, S. Dong, and E. H. Adelson, "GelSight: High-resolution robot tactile sensors for estimating geometry and force," *Sensors*, vol. 17, no. 12, 2017, Art. no. 2762.
- [31] B. Ward-Cherrier et al., "The TacTip family: Soft optical tactile sensors with 3D-printed biomimetic morphologies," *Soft Robot.*, vol. 5, no. 2, pp. 216–227, 2018.
- [32] A. Alspach, K. Hashimoto, N. Kuppaswamy, and R. Tedrake, "Soft-bubble: A highly compliant dense geometry tactile sensor for robot manipulation," in *Proc. IEEE Int. Conf. Soft Robot.*, 2019, pp. 597–604.
- [33] H. Sun and G. Martius, "Guiding the design of superresolution tactile skins with taxel value isolines theory," *Sci. Robot.*, vol. 7, no. 63, 2022, Art. no. eabm0608.
- [34] C. Trueeb, C. Sferrazza, and R. D'Andrea, "Towards vision-based robotic skins: A data-driven, multi-camera tactile sensor," in *Proc. IEEE Int. Conf. Soft Robot.*, 2020, pp. 333–338.
- [35] C. Sferrazza, A. Wahlsten, C. Trueeb, and R. D'Andrea, "Ground truth force distribution for learning-based tactile sensing: A finite element approach," *IEEE Access*, vol. 7, pp. 173438–173449, 2019.
- [36] A. Yamaguchi and C. G. Atkeson, "Recent progress in tactile sensing and sensors for robotic manipulation: Can we turn tactile sensing into vision?," *Adv. Robot.*, vol. 33, no. 14, pp. 661–673, 2019.
- [37] C. Armanini, F. Boyer, A. T. Mathew, C. Duriez, and F. Renda, "Soft robots modeling: A structured overview," *IEEE Trans. Robot.*, vol. 39, no. 3, pp. 1728–1748, Jun. 2023.

- [38] A. Yamaguchi and C. G. Atkeson, "Implementing tactile behaviors using FingerVision," in *Proc. IEEE-RAS Int. Conf. Humanoid Robot.*, 2017, pp. 241–248.
- [39] T. Kroeger, R. Timofte, D. Dai, and L. Van Gool, "Fast optical flow using dense inverse search," in *Proc. Eur. Conf. Comput. Vis.*, 2016, pp. 471–488.
- [40] C. Sferrazza and R. D'Andrea, "Design, motivation and evaluation of a full-resolution optical tactile sensor," *Sensors*, vol. 19, no. 4, 2019, Art. no. 928.
- [41] Z. Zhang, J. Dequidt, and C. Duriez, "Vision-based sensing of external forces acting on soft robots using finite element method," *IEEE Robot. Autom. Lett.*, vol. 3, no. 3, pp. 1529–1536, Jul. 2018.
- [42] M. Kaboli, K. Yao, D. Feng, and G. Cheng, "Tactile-based active object discrimination and target object search in an unknown workspace," *Auton. Robots*, vol. 43, pp. 123–152, 2019.
- [43] J. Ilonen, J. Bohg, and V. Kyrki, "Three-dimensional object reconstruction of symmetric objects by fusing visual and tactile sensing," *Int. J. Robot. Res.*, vol. 33, no. 2, pp. 321–341, 2014.
- [44] S. Wang et al., "3D shape perception from monocular vision, touch, and shape priors," in *Proc. IEEE Int. Conf. Intell. Robots Syst.*, 2018, pp. 1606–1613.
- [45] O. Yasa et al., "An overview of soft robotics," *Annu. Rev. Control, Robot., Auton. Syst.*, vol. 6, pp. 1–29, 2023.
- [46] R. Baines et al., "Multi-environment robotic transitions through adaptive morphogenesis," *Nature*, vol. 610, no. 7931, pp. 283–289, 2022.
- [47] M. Rafeeq, S. F. Toha, S. Ahmad, and M. A. Razib, "Locomotion strategies for amphibious robots—A review," *IEEE Access*, vol. 9, pp. 26323–26342, 2021.
- [48] J. Yu, R. Ding, Q. Yang, M. Tan, W. Wang, and J. Zhang, "On a bio-inspired amphibious robot capable of multimodal motion," *IEEE/ASME Trans. Mechatron.*, vol. 17, no. 5, pp. 847–856, Oct. 2012.
- [49] R. A. S. I. Subad, L. B. Cross, and K. Park, "Soft robotic hands and tactile sensors for underwater robotics," *Appl. Mechan.*, vol. 2, no. 2, pp. 356–382, 2021.
- [50] L. Li et al., "An aerial-aquatic hitchhiking robot with remora-inspired tactile sensors and thrust vectoring units," *Adv. Intell. Syst.*, 2023, Art. no. 2300381, doi: [10.1002/aisy.202300381](https://doi.org/10.1002/aisy.202300381).
- [51] A. Aggarwal, P. Kampmann, J. Lemburg, and F. Kirchner, "Haptic object recognition in underwater and deep-sea environments," *J. Field Robot.*, vol. 32, no. 1, pp. 167–185, 2015.
- [52] S. Yang et al., "Dynamic capture using a traplike soft gripper with stiffness anisotropy," *IEEE/ASME Trans. Mechatron.*, vol. 28, no. 3, pp. 1337–1346, Jun. 2023.
- [53] S. Garrido-Jurado, R. Muñoz-Salinas, F. J. Madrid-Cuevas, and M. J. Marín-Jiménez, "Automatic generation and detection of highly reliable fiducial markers under occlusion," *Pattern Recognit.*, vol. 47, no. 6, pp. 2280–2292, 2014.
- [54] C. Lanczos, *The Variational Principles of Mechanics* (Dover Books on Physics, 4th ed.). Mineola, NY, USA: Dover Publications, 1986.
- [55] A. Longva, F. Lössner, T. Kugelstadt, J. A. Fernández-Fernández, and J. Bender, "Higher-order finite elements for embedded simulation," *ACM Trans. Graph.*, vol. 39, no. 6, pp. 1–14, 2020.
- [56] N. Aigerman and Y. Lipman, "Injective and bounded distortion mappings in 3D," *ACM Trans. Graph.*, vol. 32, no. 4, pp. 1–14, 2013.
- [57] L. Liu, L. Zhang, Y. Xu, C. Gotsman, and S. J. Gortler, "A local/global approach to mesh parameterization," *Comput. Graph. Forum*, vol. 27, no. 5, pp. 1495–1504, 2008.
- [58] N. Aigerman, R. Poranne, and Y. Lipman, "Seamless surface mappings," *ACM Trans. Graph.*, vol. 34, no. 4, pp. 1–13, 2015.
- [59] J. Smith and S. Schaefer, "Bijective parameterization with free boundaries," *ACM Trans. Graph.*, vol. 34, no. 4, pp. 1–9, 2015.
- [60] W. M. Lai, D. Rubin, and E. Kreml, *Introduction to Continuum Mechanics*, 3rd ed. Amsterdam, The Netherlands: Elsevier Science, 2014.
- [61] M. Rabinovich, R. Poranne, D. Panozzo, and O. Sorkine-Hornung, "Scalable locally injective mappings," *ACM Trans. Graph.*, vol. 36, no. 2, pp. 1–16, 2017.
- [62] E. Tretschk et al., "State of the art in dense monocular non-rigid 3D reconstruction," *Comput. Graph. Forum*, vol. 42, no. 2, pp. 485–520, 2023.
- [63] M. P. Gerardo-Castro, T. Peynot, and F. Ramos, "Laser-radar data fusion with Gaussian process implicit surfaces," in *Proc. Int. Conf. Field Service Robot.*, 2015, pp. 289–302.
- [64] S. Dragiev, M. Toussaint, and M. Gienger, "Gaussian process implicit surfaces for shape estimation and grasping," in *Proc. IEEE Int. Conf. Robot. Autom.*, 2011, pp. 2845–2850.
- [65] S. Ottenhaus, M. Miller, D. Schiebener, N. Vahrenkamp, and T. Asfour, "Local implicit surface estimation for haptic exploration," in *Proc. IEEE Int. Conf. Humanoid Robots*, 2016, pp. 850–856.
- [66] C. E. Rasmussen and C. K. I. Williams, *Gaussian Processes for Machine Learning*. Cambridge, MA, USA: MIT Press, 2005.
- [67] P. Peltzer, J. Lotz, and U. Naumann, "Eigen-AD: Algorithmic differentiation of the eigen library," in *Proc. Int. Conf. Comput. Sci.*, 2020, pp. 690–704.
- [68] G. Fang, C.-D. Matte, R. B. Scharff, T.-H. Kwok, and C. C. Wang, "Kinematics of soft robots by geometric computing," *IEEE Trans. Robot.*, vol. 36, no. 4, pp. 1272–1286, Aug. 2020.
- [69] T. M. Lehmann, C. Gonner, and K. Spitzer, "Survey: Interpolation methods in medical image processing," *IEEE Trans. Med. Imag.*, vol. 18, no. 11, pp. 1049–1075, Nov. 1999.
- [70] B. G. Kitchener, J. Wainwright, and A. J. Parsons, "A review of the principles of turbidity measurement," *Prog. Phys. Geogr.*, vol. 41, no. 5, pp. 620–642, 2017.
- [71] H. Lee, D. Jeong, H. Yu, and J. Ryu, "Autonomous underwater vehicle control for fishnet inspection in turbid water environments," *Int. J. Control, Auton. Syst.*, vol. 20, no. 10, pp. 3383–3392, 2022.
- [72] J. Li, C. Huang, Y. Zha, C. Wang, N. Shang, and W. Hao, "Spatial variation characteristics and remote sensing retrieval of total suspended matter in surface water of the Yangtze river," *Environ. Sci.*, vol. 42, no. 12, pp. 5239–5249, 2021.
- [73] N. Guo et al., "Autoencoding a soft touch to learn grasping from on-land to underwater," *Adv. Intell. Syst.*, vol. 6, no. 1, 2024, Art. no. 2300382.
- [74] A. Thayananthan, B. Stenger, P. Torr, and R. Cipolla, "Shape context and chamfer matching in cluttered scenes," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, 2003, pp. I-1, doi: [10.1109/CVPR.2003.1211346](https://doi.org/10.1109/CVPR.2003.1211346).
- [75] H. Ravichandar, A. S. Polydoros, S. Chernova, and A. Billard, "Recent advances in robot learning from demonstration," *Annu. Rev. Control, Robot., Auton. Syst.*, vol. 3, pp. 297–330, 2020.
- [76] C. Chi et al., "Diffusion policy: Visuomotor policy learning via action diffusion," in *Proc. Robot. Sci. Syst.*, 2023. [Online]. Available: <https://roboticsconference.org/2023/program/papers/026/>
- [77] T. Wu et al., "Vision-based tactile intelligence with soft robotic metamaterial," *Mater. Des.*, vol. 238, 2024, Art. no. 112629.
- [78] F. Wan and C. Song, "SeeThruFinger: See and grasp anything with a soft touch," 2023, *arXiv:2312.09822*.
- [79] G. Li et al., "Self-powered soft robot in the Mariana trench," *Nature*, vol. 591, no. 7848, pp. 66–71, 2021.
- [80] J. F. Esmail, M. Z. Mohamedmeki, and A. E. Ajeel, "Using the uniaxial tension test to satisfy the hyperelastic material simulation in ABAQUS," in *Proc. IOP Conf. Ser.: Mater. Sci. Eng.*, 2020, vol. 888, Art. no. 012065.



**Ning Guo** (Student Member, IEEE) received the B.E. degree in mechanical design, fabrication, and automation, and the M.S. degree in mechatronics engineering from the Harbin Institute of Technology, Harbin, China, in 2015 and 2017, respectively. He is currently working toward the Ph.D. degree in mechanics with the Department of Mechanical and Energy Engineering, Southern University of Science and Technology, Shenzhen, China.

His research interests include soft robotics and robot learning.



**Xudong Han** (Student Member, IEEE) received the B.E. degree in mechanical engineering in 2022 from the Southern University of Science and Technology, Shenzhen, China, where he is currently working toward the Ph.D. degree in mechanics with the Department of Mechanical and Energy Engineering.

His research interests include soft robotics and robot learning.





**Shuqiao Zhong** received the B.E. degree in industrial design from the Harbin Institute of Technology, Harbin, China, in 2019. He is currently working toward the Ph.D. degree in mechanics with the Department of Ocean Science and Engineering, Southern University of Science and Technology, Shenzhen, China.

His research interests include underwater grasping and gripper design.



**Zhiyuan Zhou** received the bachelor's degree in geophysics and Ph.D. degree in solid earth physics from Tongji University, Shanghai, China, in 2009 and 2015, respectively.

He was a Visiting Student and Scholar with the Woods Hole Oceanographic Institution (WHOI). He was an Associate Scientist with the South China Sea Institute of Oceanology, Chinese Academy of Sciences, from 2015 to 2021. He is a Research Associate Professor with the Department of Ocean Science and Engineering, SUSTech, Shenzhen, China. His

research interests include marine geophysics and ocean artificial intelligence.

Dr. Zhou was awarded a grand prize in the Ocean Science and Technology Award in China in 2022 (6/20). He is an Editorial Board Member of "Earth Science and Engineering."



**Jian Lin** received the bachelor's degree in geophysics from the University of Science and Technology of China, Hefei, China, in 1982 and the Ph.D. degree in earth sciences from Brown University, Providence, RI, USA, in 1988.

He was a Senior Scientist with the Woods Hole Oceanographic Institution (WHOI) and a Faculty Member with the Massachusetts Institute of Technology (MIT)/WHOI Joint Program in Oceanography. He is currently a Chair Professor with the Department of Ocean Science and Engineering, SUSTech,

Shenzhen, China. His research interests include marine geophysics and ocean technology.

Dr. Lin was elected as a Member of Academia Europaea, a Member of the European Academy of Sciences, a Member of the Russian Academy of Natural Sciences, and a Foreign Fellow of the Pakistan Academy of Sciences. He was also elected as a Fellow of the American Geophysical Union (AGU), a Fellow of the American Association for the Advancement of Science (AAAS), and a Fellow of the Geological Society of America (GSA). He received the Henry Bigelow Chair Award for Excellence in Oceanography.



**Jian S. Dai** (Fellow, IEEE) received the Ph.D. degree in advanced kinematics, mechanisms, and robotics from the University of Salford, Salford, U.K., in 1993.

He is a Chair Professor with the Department of Mechanical and Energy Engineering and the Director of the SUSTech Institute of Robotics, Southern University of Science and Technology, Shenzhen, China. He has worked in mechanisms and robotics for the past 30 years and has authored or coauthored more than 600 peer-reviewed papers. His research interests

include theoretical and computational kinematics, reconfigurable mechanisms, dexterous manipulators, end effectors, and multifingered hands.

Dr. Dai is a Fellow of the Royal Academy of Engineering, the American Society of Mechanical Engineers (ASME), and the Institution of Mechanical Engineers. He received the Mechanisms and Robotics Award, given annually by the ASME Mechanisms and Robotics Committee to engineers known for a lifelong contribution to the mechanism design or theory field, as the 27th recipient since 1974. He also received the 2020 ASME Machine Design Award, the 58th recipient since the Award was established in 1958. He was the recipient of several journal and conference best paper awards. He is the Editor-in-Chief for *Robotica*, an Associate Editor for the *ASME Transactions: Journal of Mechanisms and Robotics*, and a Subject Editor for *Mechanism and Machine Theory*.



**Fang Wan** (Member, IEEE) received the bachelor's degree in applied mathematics from the Huazhong University of Science and Technology, Wuhan, China, in 2009, and the doctoral degree in mathematics and physics from Nanyang Technological University, Singapore, in 2014.

She is currently an Assistant Professor with the School of Design, Southern University of Science and Technology, Shenzhen, China. Her research interests include applied mathematical tools and data-driven design technologies.

Dr. Wan is currently a Member of the Design Society.



**Chaoyang Song** (Senior Member, IEEE) received the B.Eng. degree in mechanical engineering from Tongji University, Shanghai, China, in 2009 and the Ph.D. degree in mechanism and robotics from Nanyang Technological University, Singapore, in 2014.

From 2013 to 2015, he was a Postdoctoral Research Fellow with the Massachusetts Institute of Technology and the Singapore University of Technology and Design. Then, he was a Lecturer (Assistant Professor) with the Department of Mechanical and Aerospace Engineering, Monash University. He is currently an

Assistant Professor with the Department of Mechanical and Energy Engineering and an Adjunct Assistant Professor with the Department of Computer Science and Engineering, Southern University of Science and Technology, Shenzhen, China. His research interest includes the design science of bionic robotics and robot learning.