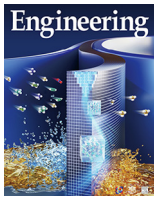
Engineering xxx (xxxx) xxx

Contents lists available at ScienceDirect

Engineering

journal homepage: www.elsevier.com/locate/eng



Research
Artificial Intelligence–Article

A Reconfigurable Data Glove for Reconstructing Physical and Virtual Grasps

Hangxin Liu [a,*,†], Zeyu Zhang [a,b,†], Ziyuan Jiao [a,b,†], Zhenliang Zhang [a], Minchen Li [c], Chenfanfu Jiang [c], Yixin Zhu [d,*], Song-Chun Zhu [a,d,e]

[a] National Key Laboratory of General Artificial Intelligence, Beijing Institute for General Artificial Intelligence (BIGAI), Beijing 100080, China
[b] Center for Vision, Cognition, Learning, and Autonomy, University of California, Los Angeles, CA 90095, USA
[c] Multi-Physics Lagrangian-Eulerian Simulations Lab, Department of Mathematics, University of California, Los Angeles, CA 90095, USA
[d] Institute for Artificial Intelligence, Peking University, Beijing 100871, China
[e] Department of Automation, Tsinghua University, Beijing 100084, China

ARTICLE INFO

Article history:
Received 8 March 2022
Revised 23 December 2022
Accepted 4 January 2023

Keywords:
Data glove
Tactile sensing
Virtual reality
Physics-based simulation

ABSTRACT

In this work, we present a reconfigurable data glove design to capture different modes of human hand–object interactions, which are critical in training embodied artificial intelligence (AI) agents for fine manipulation tasks. To achieve various downstream tasks with distinct features, our reconfigurable data glove operates in three modes sharing a unified backbone design that reconstructs hand gestures in real time. In the tactile-sensing mode, the glove system aggregates manipulation force via customized force sensors made from a soft and thin piezoresistive material; this design minimizes interference during complex hand movements. The virtual reality (VR) mode enables real-time interaction in a physically plausible fashion: A caging-based approach is devised to determine stable grasps by detecting collision events. Leveraging a state-of-the-art finite element method, the simulation mode collects data on fine-grained four-dimensional manipulation events comprising hand and object motions in three-dimensional space and how the object's physical properties (e.g., stress and energy) change in accordance with manipulation over time. Notably, the glove system presented here is the first to use high-fidelity simulation to investigate the unobservable physical and causal factors behind manipulation actions. In a series of experiments, we characterize our data glove in terms of individual sensors and the overall system. More specifically, we evaluate the system's three modes by ① recording hand gestures and associated forces, ② improving manipulation fluency in virtual reality (VR), and ③ producing realistic simulation effects of various tool uses, respectively. Based on these three modes, our reconfigurable data glove collects and reconstructs fine-grained human grasp data in both physical and virtual environments, thereby opening up new avenues for the learning of manipulation skills for embodied AI agents.

1. Challenges in learning manipulation

Manipulation and grasping are among the most fundamental topics in robotics. This classic field has been rejuvenated by the recent boom in embodied artificial intelligence (AI), wherein an agent (e.g., a robot) is tasked to learn by interacting with its environment. Since then, learning-based methods have been widely applied and have elevated robots' manipulation competence. Often, robots either train on data directly obtained from sensors (e.g., object grasping from a cluster [1,2], pick-and-place [3], object handover [4], or door opening [5]) or learn from human demonstrations (e.g., motor motions [6,7], affordance [8,9], task structure [10–12], or reward functions [13–15]).

Learning meaningful manipulation has a unique prerequisite: It must incorporate fine-grained physics to convey an understanding of the complex process that occurs during the interaction. Although we have witnessed the solid advancement of certain embodied AI tasks (e.g., visual-language navigation), these successes are primarily attributed to the readily available plain images and their annotations (pixels, segments, or bounding boxes) that

* Corresponding authors.
E-mail addresses: liuhx@bigai.ai (H. Liu), yixin.zhu@pku.edu.cn (Y. Zhu).

† These authors contributed equally to this work.

https://doi.org/10.1016/j.eng.2023.01.009

Please cite this article as: H. Liu, Z. Zhang, Z. Jiao et al., A Reconfigurable Data Glove for Reconstructing Physical and Virtual Grasps, Engineering, https://doi.org/10.1016/j.eng.2023.01.009

are directly extracted from the existing training platforms [16–18], while physics information during the interactions is still lacking. Similarly, although modern vision-based sensors and motion-capture systems can collect precise trajectory information, neither can precisely estimate physical properties during interactions. Existing software and hardware systems are insufficient for learning sophisticated manipulation skills for the following three reasons:

First, understanding fine-grained manipulation or human–object interactions requires a joint understanding of both hand gesture[1] and force [20]; distinguishing certain actions purely based on the hand gesture is challenging, if not impossible. For example, in the task of opening a medicine bottle that requires either pushing or squeezing the lid to unlock the childproof mechanism, it is insufficient to differentiate the opening actions by visual information alone, because the pushing and squeezing actions are visually similar (or even identical) to each other [21]. Reconstructing hand gestures or trajectories alone has already been shown to be challenging, as severe hand–object occlusion hinders the data collection reliability. To tackle this problem, we introduce a tactile-sensing glove to jointly capture hand gestures through a network of inertial measurement units (IMUs) and force exerted by the hand using six customized force sensors during manipulation. The force sensors are constructed from Velostat—a piezoresistive fabric with changing resistance under pressures, which is soft and thin to allow natural hand motions. Together, the force sensors provide a holistic view of manipulation events. A preliminary version of this system has been presented in the work of Liu et al. [20] (Appendix A).

Second, contact points between hand and object play a significant role in understanding why and how a specific grasp is chosen. Such information is traditionally challenging to obtain (e.g., through thermal imaging [22]). To address this challenge, we devise a VR glove and leverage VR platforms to obtain contact points. This design incorporates a caging-based approach to determine a stable grasp of a virtual object based on the collision geometry between fingers and the object. The collisions trigger a network of vibration motors on the glove to provide haptic feedback. The VR glove jointly collects trajectory and contact information that is otherwise difficult to obtain physically. A preliminary version of this system has been presented in the work of Liu et al. [23] (Appendix A).

Third, much attention has been paid to collecting hand information during fine manipulation but not to the object being manipulated or its effects caused by actions. This deficiency prohibits the use of collected data for studying complex manipulation events. For example, consider a tool-use scenario. A manipulation event cannot be comprehensively understood without capturing the interplay between the human hand, the tool being manipulated, and the action effects. As such, this perspective demands a solution beyond the classic hand-centric view in developing data gloves. Furthermore, since the effects caused by the manipulation actions are traditionally difficult to capture, they are often treated as a task of recognizing discrete, symbolic states or attributes in computer vision [24–26], losing their intrinsic continuous nature. To overcome these limits of traditional data gloves, we propose to integrate a physics-based simulation using the state-of-the-art finite element method (FEM) [19] to model object fluents—the time-varying states in the event [27]—and other physical properties involved, such as contact forces and the stress within the object. This glove with simulation captures a human manipulation action and analyzes it in four-dimensional (4D) space by including: ① the contact and geometric information of the hand gesture and the object in three-dimensional (3D) space, and ② the transition and coherence between the object's fluent changes and the manipulation events over time. To the best of our knowledge, this is the first time such 4D data offering a holistic view of manipulation events is used in this field, and its use will open up new avenues for studying manipulations and grasping.

Sharing a unified backbone design that reconstructs hand gestures in real-time, the proposed data glove can be easily reconfigured to ① capture force exerted by hand using piezoresistive material, ② record contact information by grasping stably in VR, or ③ reconstruct both visual and physical effects during the manipulation by integrating physics-based simulation. Our system extends the long history of developing data gloves [28] and endows embodied AI agents with a deeper understanding of hand–object interactions.

This paper makes three contributions compared with prior work [20,23]. First, we introduce the concept of a reconfigurable glove-based system. The three operating modes tackle a broader range of downstream tasks with distinct features. This extension does not sacrifice the easy-to-replicate nature, as different modes share a unified backbone design. Second, a state-of-the-art FEM-based physical simulation is integrated to augment the grasp data with simulated action effects, thereby providing new opportunities for studying hand–object interactions and complex manipulation events. Third, we demonstrate that the data collected by our glove-based system—either virtually or physically—is effective for learning in a series of case studies.

1.1. Related work

1.1.1. Hand gesture sensing

Recording finger joints' movements is the core of hand gesture sensing. Various types of hardware have been adopted to acquire hand gestures. Although curvature/flex sensors [29,30], liquid metal [31], a stretchable strain sensor [32], and triboelectric material [33] are among proven approaches, these can only measure unidirectional bending angles. Hence, they are less efficient for recording a hand's metacarpophalangeal (MCP) joints with two degrees of freedom (DoFs) for finger abduction and adduction. In addition, by wrapping around bending finger joints, these instruments sacrifice natural hand movements due to their large footprint and rigidness. In comparison, IMUs can measure one phalanx's 6-DoF pose, interfere less with joint motions, and perform more consistently over an extended period of time. As a result, adopting IMUs in data gloves has prevailed in modern design, including IMUs channeled by a Zigbee network [34], a circuit board with a 6-DoF accelerometer/gyroscope and a 3-DoF magnetometer placed on each of the 15 phalanxes [35], and a population of IMUs connected through flexible cables [36]. Often, the raw sensory information requires further filtering [37] and estimation [35,38,39].

1.1.2. Force sensing

Sensing the forces exerted by a hand during manipulation has attracted growing research attention and requires a more integrated glove-based system. Here, we highlight some signature designs. An elastomer sensor with embedded liquid–metal material [40] was able to sense force across a large area (e.g., the palm) and estimate joint movements by measuring skin strain. FlexiForce sensors can acquire hand forces [41], while an optical-based motion-capture system tracks hand gestures. Forces and gestures can also be estimated using 9-DoF IMUs without additional hardware [42], although the force estimation is crude. Other notable designs involve specialized hardware, including force-sensitive resistors [43] and a specific tactile sensor for fingertips [44]. Recently, soft films made from piezoresistive materials whose

[1] In this article, the phrase "hand gesture" is used to refer to the collective movement of the fingers and palm, whereas "hand pose" is used to refer to the position and orientation of the wrist.

resistance changes under pressing forces (e.g., Velostat) have become increasingly popular in robotic applications; this type of material permits force sensing without constraining the robots' or human hand's motions [45–48].

1.2. Overview: The three modes of the reconfigurable data glove

To tackle the aforementioned challenges and fill in the gap in the literature, we devised a reconfigurable data glove that is capable of operating in three modes for various downstream tasks with distinct features and goals.

1.2.1. Tactile-sensing mode

We start with a glove design using an IMU configuration [35] to reconstruct hand gestures. Our system's software and hardware designs are publicly available for easy replication. A customized force sensor made from Velostat—a soft fabric whose resistance changes under different pressures—is adopted to acquire the force distributions over large areas of the hand without constraining natural hand motions. Fig. 1(a) [19,20,23] summarizes this tactile-sensing glove design.

1.2.2. VR mode

By reconstructing virtual grasps in VR, this mode provides supplementary contact information (e.g., contact points on an object) during manipulation actions. In contrast to the dominating symbolic grasp methods that directly attach the virtual object to the virtual hand when a grasp event is triggered [49], our glove-based system enables a natural and realistic grasp experience with a fine-grained hand gesture reconstruction and force estimated at specific contact points; a symbolic grasp would cause finger penetrations or non-contacting, since the attachments between the hand and object are predefined. Although collecting grasp-related data in VR is more convenient and economical than other specialized data-acquisition pipelines, the lack of direct contact between the hand and physical objects inevitably leads to less natural interactions. Thus, providing haptic feedback is critical to compensate for this drawback. We use vibration motors to provide generic haptic feedback to each finger, thereby increasing the realism of grasping in VR. Fig. 1(b) [19,20,23] summarizes the VR glove design.

1.2.3. Simulation mode

Physics-based simulations emulate a system's precise changes over time, thus opening up new directions for robot learning [50], including learning robot navigation [16], bridging human and robot embodiments in learning from demonstration [12], soft robot locomotion [51], liquid pouring [52], and robot cutting [53]. In a similar vein, simulating how an object's fluent changes as the result of a given manipulation action provides a new perspective on hand–object interactions. In this article, we adopt a state-of-the-art FEM simulator [19] to emulate the causes and effects of manipulation events. As shown in Fig. 1(c) [19,20,23], by integrating physical data collected by the data glove with simulated effects, our system reconstructs a new type of 4D manipulation data with high-fidelity visual and physical properties on a large scale. We believe that this new type of data can significantly impact how manipulation datasets are collected in the future and can assist in a wide range of manipulation tasks in robot learning.

1.3. Structure of this article

The remainder of this article is organized as follows. We start with a unified design for hand gesture sensing in Section 2. With different goals, the tactile-sensing mode [20] and the VR mode [23] are presented in Section 3 and Section 4, respectively. A new state-of-the-art, physics-based simulation using FEM [54] is integrated in Section 5 to collect 4D manipulation data, which is the very first in the field to achieve such high fidelity, to the best of our knowledge. We evaluate our system in three modes in Section 6 and conclude the paper in Section 7.

2. A unified backbone design for gesture sensing

This section introduces the IMU setup for capturing hand gestures in Section 2.1. As this setup is shared among all three modes of the proposed reconfigurable data glove, we further evaluate the IMU performance in Section 2.2.

2.1. Hand gesture reconstruction

2.1.1. IMU specification

Fifteen Bosch BNO055 9-DoF IMUs are deployed for hand gesture sensing. One IMU is mounted to the palm, two IMUs to the thumb's distal and intermediate phalanges, and the remaining 12 are placed on the phalanxes of the other four fingers. Each IMU includes a 16-bit triaxial gyroscope, a 12-bit triaxial accelerometer, and a triaxial geomagnetometer. This IMU is integrated with a built-in proprietary sensor fusion algorithm running on a 32-bit microcontroller, yielding each phalanx's pose in terms of a quaternion. The geomagnetometer acquires an IMU's reference frame to the Earth's magnetic field, supporting the pose calibration protocol (introduced later). The small footprint of the BNO055 (5.0 cm $\times$ 4.5 cm) allows easy attachment to the glove and minimizes interference with natural hand motions. A pair of TCA9548A I^2C multiplexers is used for networking the 15 IMUs and connecting them to the I^2C bus interfaces on a Raspberry Pi 2 Model B board (henceforth RPi for brevity); RPi acts as the master controller for the entire glove system.

2.1.2. Hand forward kinematics

A human hand has about 20 DoFs: both the proximal interphalangeal (PIP) joint and the distal interphalangeal (DIP) joint have one DoF, whereas an MCP joint has two. Based on this anatomical structure, we model each finger by a 4-DoF kinematic chain whose base frame is the palm and the end-effector frame is the distal phalanx. The thumb is modeled as a 3-DoF kinematic chain consisting of a DIP joint and an MCP joint.

After obtaining a joint's rotational angle using two consecutive IMUs, the position and orientation of each phalanx can be computed by forward kinematics. Fig. 2 [20] shows an example of the index finger's kinematic chain and the attached frame. Frame 1 is assigned to the palm, and Frames 2, 3, and 4 are assigned to the proximal, middle, and distal phalanx, respectively. The proximal, middle, and distal phalanx lengths are respectively denoted by l_1, l_2, and l_3. The flexion and extension angles of the MCP, PIP, and DIP joints are denoted as θ_1, θ_2, and θ_3, respectively. In addition, the MCP joint has an abduction and adduction angle denoted as β. d_x and d_y are the offsets in the x and y directions between the palm's center and the MCP joint. Table 1 derives the Denavit–Hartenberg (D–H) parameters for each reference frame, wherein a general homogeneous transformation matrix $\boldsymbol{T}$ from frames $i-1$ to i (where i is the Frame index mentioned above) can be given by the following:

$$ {}^{i-1}_{i}\boldsymbol{T}_{i-1} = \begin{bmatrix} \cos\theta_i & -\sin\theta_i & 0 & a_{i-1} \\ \sin\theta_i\cos\alpha_{i-1} & \cos\theta_i\cos\alpha_{i-1} & -\sin\alpha_{i-1} & -\sin\alpha_{i-1}d_i \\ \sin\theta_i\sin\alpha_{i-1} & \cos\theta_i\sin\alpha_{i-1} & \cos\alpha_{i-1} & \cos\alpha_{i-1}d_i \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{1} $$

where α_{i-1}, a_{i-1}, θ_{i-1}, and d_i are the D–H parameters.

Fig. 1. Overview of our reconfigurable data glove in three operating modes, which share a unified backbone design of an IMU network that captures the hand gesture. (a) The tactile-sensing mode records the force exerted by the hand during manipulation [20]. (b) The VR mode supports stable grasping of virtual objects in VR applications and provides haptic feedback via vibration motors [23]. Contact configurations are conveniently logged. (c) The simulation mode incorporates state-of-the-art FEM simulation [19] to augment the grasp data with fine-grained changes in the object's properties. *N*: the number of unit(s) used in the prototype.

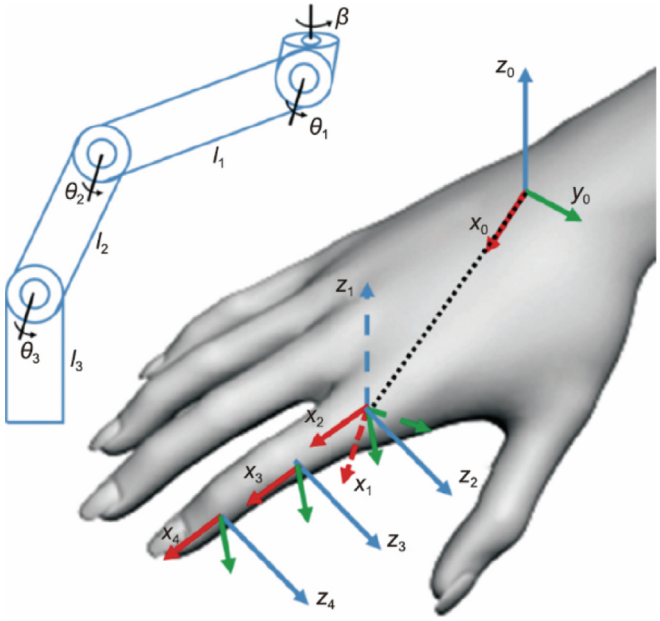


Fig. 2. The kinematic chain of the index finger with coordinate frames attached.β: the abduction and adduction angle of MCP joint; l_1, l_2, and l_3: the proximal, middle, and distal phalanx lengths; θ_1, θ_2, and θ_3: the flexion and extension angles of the MCP, PIP, and DIP joints; $x_{(\cdot)}$, $y_{(\cdot)}$, and $z_{(\cdot)}$: the *x*, *y*, and *z* coordinate attached to the corresponding frame. Reproduced from Ref. [20] with permission.

Table 2 lists the homogeneous transformation matrices of each phalanx, which can be used to express each phalanx's pose in the palm's reference frame in the cartesian space. The forward kinematics model keeps better track of the sensed hand gesture by reducing the inconsistency due to IMU fabrication error and anatomical variations among the users' hands.

2.1.3. Joint limits

We adopt a set of commonly used inequality constraints [55] to limit the motion ranges of the finger joints, thereby eliminating unnatural hand gestures due to sensor noise:

$$\begin{aligned} &\text{MCP joint}: \begin{cases} 0^\circ \le \theta_1 \le 90^\circ \\ -15^\circ \le \beta \le 15^\circ \end{cases} \\ &\text{PIP joint}: 0^\circ \le \theta_2 \le 110^\circ \\ &\text{DIP joint}: 0^\circ \le \theta_3 \le 90^\circ \end{aligned} \tag{2}$$

2.1.4. Pose calibration

Inertial sensors such as IMUs suffer from a common problem of drifting, which causes an accumulation of errors during operations. To overcome this issue, we introduce an IMU calibration protocol. When the sensed hand gesture degrades significantly, the user wearing the glove can hold the hand flat and maintain this gesture (Fig. 3 [23]) to initiate calibration; the system records the relative pose between the IMU and world frames. The orientation data measured by the IMUs are multiplied by the inverse of this relative pose to cancel out the differences, thus eliminating accumulated errors due to drifting. This routine can be performed conveniently when experiencing unreliable hand gesture sensing results.

2.2. IMU evaluation

We evaluated an individual IMU's bias and variance during rotations. Furthermore, we examined how accurately two articulated IMUs can reconstruct a static angle, indicating the performance of an atomic element in sensing the finger joint angle.

2.2.1. Evaluations of a single IMU

As the reliability of the gesture sensing primarily depends on the IMU performance, it is crucial to investigate the IMU's bias and variance. More specifically, we rotated an IMU using a precise

Table 1
Denavit–Hartenberg parameters of a finger.

Link Index	Parameter			
	α_{i-1}	a_{i-1}	θ_i	d_i
1	0	0	β	0
2	$\pi/2$	l_1	θ_1	0
3	0	l_2	θ_2	0
4	0	l_3	θ_3	0

Table 2
Concatenation of transformation matrices.

Phalanx	Transformation
Proximal	${}^0_1T\,{}^1_0T\,{}^1_1T$
Middle/distal for thumb	${}^0_1T\,{}^1_0T\,{}^1_1T\,{}^2_3T\,{}^2_2T$
Distal	${}^0_1T\,{}^1_0T\,{}^1_1T\,{}^2_3T\,{}^2_2T\,{}^3_4T\,{}^3_3T$

stepper motor controlled by an Arduino microcontroller. Four rotation angles—90°, 180°, 270°, and 360°—were executed 20 times each at a constant angular velocity of 60 r·min^{-1}. We did not test for a rotation angle exceeding 360°, as this is beyond the fingers' motion range. Fig. 4(a) [20] summarizes the mean and the standard deviation of the measured angular error. Overall, the IMU performed consistently with a bias between 2° and 3° and a ±1.7° standard deviation, suggesting that post-processing could effectively reduce the sensor bias.

2.2.2. Evaluations of articulated IMUs

Evaluating IMU performance on whole-hand gesture sensing is difficult due to the lack of ground truth. As a compromise, we 3D printed four rigid bends with angles of 0°, 45°, 90°, and 135° to emulate four specific states of finger bending, which evenly divided a finger joint's motion range as defined in Eq. (2). Using two IMUs to construct a bend, assuming it to be a revolute joint, we tested the accuracy of the reconstructed joint angle by computing the relative poses between the two IMUs. Fig. 4(b) [20] shows the errors of the estimated joint angles. Fig. 4(c) [20] shows a schematic of this experimental setup, and Fig. 4(d) [20] shows the physical setup with a 90° bending angle. During the test, one IMU was placed 2 cm behind the bend, and another was placed 1 cm ahead, simulating the IMUs attached to a proximal phalanx and a middle phalanx, respectively. We repeated the test 20 times for each rigid bend. As the bending angle increased, the reconstruction errors increased from 4° to about 6°, with a slightly expanded confidence interval. Overall, the errors were still reasonable, although the IMUs tended to underperform as the bending angle increased. Through combination with the pose calibration protocol, these errors can be better counterbalanced, and the utilized IMU network can reliably support the collection of grasping data (see Section 6 for various case studies).

3. Tactile-sensing mode

Our reconfigurable data glove can be easily configured to the tactile-sensing mode, which shares the unified backbone design described in Section 2. The tactile-sensing mode measures the distribution of forces exerted by the hand during complex hand–object interactions. We start by describing the force sensor specifications in Section 3.1, which is followed by details of prototyping in Section 3.2. We conclude this section with a qualitative evaluation in Section 3.3.

3.1. Force sensor

We adopt a network of force sensors made from Velostat to provide force sensing in this tactile-sensing mode. Fig. 5(a) [20] illustrates the Velostat force sensor's multi-layer structure. A taxel (i.e., a single-point force-sensing unit) is composed of one inner layer of Velostat (2 cm × 2 cm) and two middle layers of conductive fabric, stitched together by conductive thread and enclosed by two outer layers of insulated fabric. A force-sensor pad consisting of two taxels is placed on each finger, and a sensor grid with 4 × 4 taxels is placed on the palm. Lead wires to the pads and grid are braided into the conductive thread.

As the Velostat's resistance changes with different pressing forces, the measured voltage across a taxel can be regarded as the force reading at that region. To acquire the voltage readings, we connect these Velostat force-sensing taxels in parallel via analog multiplexers controlled by the RPi's GPIO and output to its serial peripheral interface (SPI)-enabled ADS1256 analog to digital converter (ADC). More specifically, two 74HC4051 multiplexers are

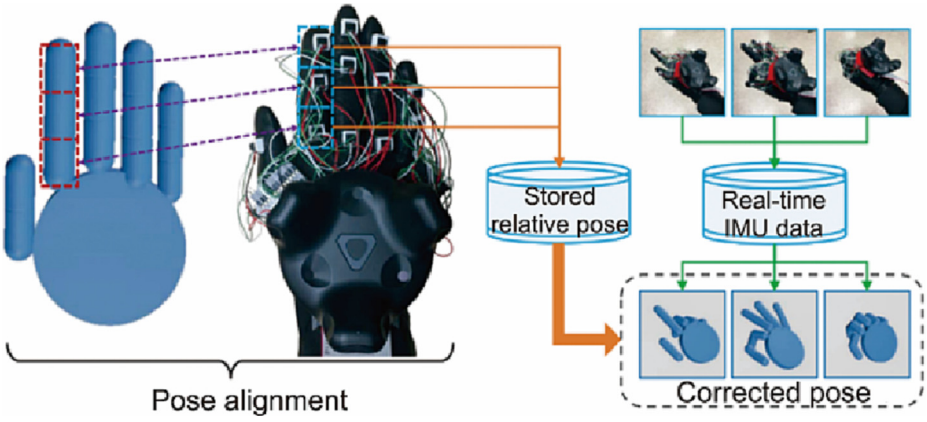


Fig. 3. The IMU calibration protocol. The protocol starts by holding the hand flat, as shown by the virtual hand model. The relative pose between the world frame and the IMU's local coordinate system is recorded. The inverse of the recorded relative pose corrects the IMU data. Reproduced from Ref. [23] with permission.

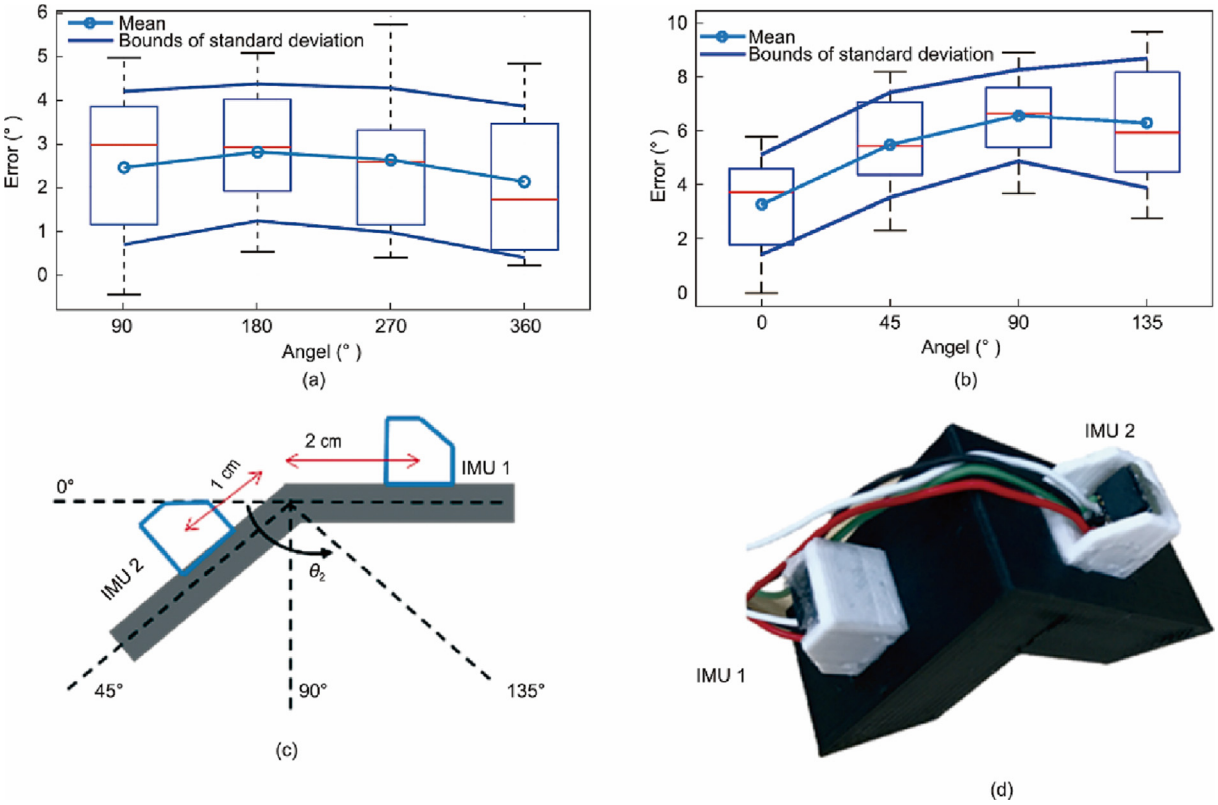


Fig. 4. Evaluations of IMU performance. The measurement error is summarized as the mean and standard deviation of (a) a single IMU and (b) two articulated IMUs under different settings. The red horizontal lines, blue boxes, and whiskers indicate the median error, the 25th and 75th percentiles, and the range of data points not considered to be outliers, respectively. (c) A schematic of the experimental setup for evaluating the angle reconstruction with two articulated IMUs is and (d) its physical setup with a 90° bending angle. Reproduced from Ref. [20] with permission.

used for the palm grid, and a CD74HC4067 multiplexer is used for all the finger pads. A voltage divider circuit, shown in Fig. 5(b) [20], is constructed by connecting a 200 Ω resistor between the RPi's ADC input channel and the multiplexers.

We now characterize the sensor's force–voltage relation [56]. A total of 13 standard weights (0.1–1.0 kg with 0.1 kg increments, 1.2, 1.5, and 2.0 kg) were applied to a taxel, and the associated voltages across that taxel were measured. The calibration circuit was the same as that in Fig. 5(b) [20], except that only the taxel of interest was connected. The weights in kilograms were converted to forces in Newtons with a gravitational acceleration $g = 10\ \mathrm{m{\cdot}s^{-2}}$. We first tested the power law [56] for characterizing the force–voltage relation of a taxel. The result was $F = -1.067\ V^{-0.4798} + 3.244$, with the correlation coefficient $R^2 = 0.9704$, where F is the applied force, and V is the output voltage. However, we further tested a logarithmic law, resulting in a better force–voltage relation: $F = 0.569 \times \log(44.98\ V)$ with a higher $R^2 = 0.9902$. Hence, we adopted the logarithmic fit to establish a correspondence between the voltage reading across a taxel and the force the taxel is subjected to. Fig. 5(c) [20] compares these two fits.

3.2. Prototyping

Fig. 1(a) [19,20,23] displays a prototype of the tactile-sensing glove. The capability of force sensing is accomplished by placing one Velostat force-sensing pad on each finger (one taxel in the proximal area and another in the distal area) and a single 4 × 4 Velostat force-sensing grid over the glove's palm region. Based on the established force–voltage relation, these taxels collectively measure the distribution of forces exerted by the hand. Meanwhile, the 15 IMUs capture the hand gestures in motion. These components are all connected to the RPi, which can be remotely accessed to visualize and subsequently utilize the collected gesture and force data in a local workstation, providing a neat solution to collect human manipulation data.

By measuring the voltage and current across each component, we investigated the power consumption of the prototype. Table 3 reports the peak power of each component of interest as the product of its voltage and current in a ten-min operation. The total power consumption was 2.72 W, which can be easily powered by a conventional Li-Po battery, offering an untethered user experience and natural interactions during data collection.

3.3. Qualitative evaluation

We evaluated the performance of the tactile-sensing glove in differentiating among low, medium, and high forces by grasping a water bottle in three states, empty, half-full, and full, whose weights were 0.13, 0.46, and 0.75 kg, respectively. The participants were asked to perform the grasps naturally and succinctly—exerting a force just enough to prevent the bottle from slipping out of the hand; Fig. 5(d) [20] shows such an instance. Ten grasps were performed for each bottle state. To simplify the analysis, the force in the palm was the average of all 16 force readings of the palm grid, and the force in each finger was the average reading of the corresponding finger pads. Fig. 5(e) [20] shows the recorded forces exerted by different hand regions.

4. VR mode

Since the different modes of our data glove share a unified backbone design, reconfiguring the glove to the VR mode in order to

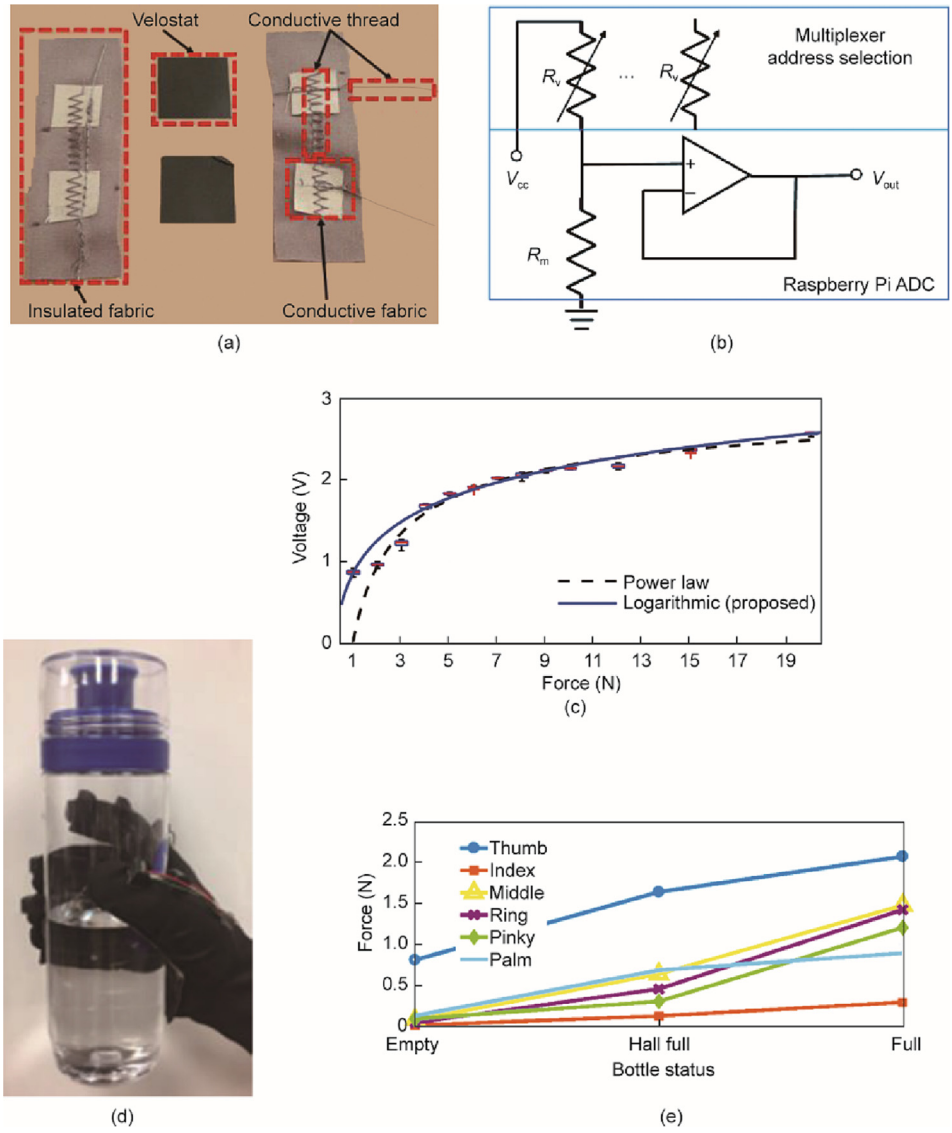


Fig. 5. Characterization of the Velostat force sensor. (a) The multi-layer structure of a Velostat force sensor. (b) The circuit layout for force data acquisition. (c) The force–voltage relation of one sensing taxel. Instead of using a power law, our choice of a logarithmic law fits the data better. (d) A grasp of the half-full bottle. (e) Force responses of grasping empty, half-full, and full bottles, respectively. ADC: analog to digital converter. R_v: the resistance of one Velostat taxel; R_m: the resistance of the voltage divider; V_{cc}: the input voltage; V_{out}: the measured voltage. Reproduced from Ref. [20] with permission.

Table 3
Power consumption of the tactile-sensing glove.

Component	Power (W)
Gesture sensing	
15 IMUs	0.60
Force sensing	
6 Velostat	0.02
Computing	
RPi	2.15
Total	2.77

obtain contact points during interactions can be achieved with only three steps. First, given the sensed hand gestures obtained by the shared backbone, we need to construct a virtual hand model for interactions (see Section 4.1). Next, we must develop an approach to achieve a stable grasp of virtual objects (see Section 4.2). Finally, grasping objects in VR introduces new difficulty without a tangible object being physically manipulated; we leverage haptic feedback to address this problem in Section 4.3. We conclude this section with an evaluation in Section 4.4.

4.1. Virtual hand model

Generating a stable grasp is the prerequisite for obtaining contact points during interactions. Existing vision-based hand gesture sensing solutions, including commercial projects such as LeapMotion [57] and RealSense [58], struggle with stable grasps due to occlusions, sensor noises, and a limited field of view (FoV); interested readers can refer to Fig. 6(a) [23] for a comparison in a typical scenario. In comparison, existing VR controllers adopt an alternative approach—the virtual objects are directly attached to the virtual hand when a grasp event is triggered. As illustrated in Fig. 6 (b) [23], the resulting experience has minimal realism and cannot reflect the actual contact configuration. The above limitations motivate us to realize a stable virtual grasp by developing a caging-based approach that is capable of real-time computation

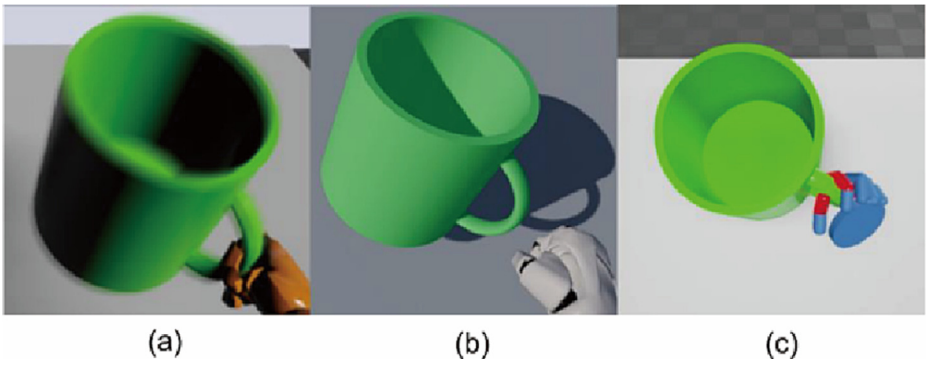


Fig. 6. Comparison of a grasp among (a) a LeapMotion sensor, (b) an Oculus Touch controller, and (c) our reconfigurable glove system in the VR mode. The grasp in (a) is unstable, as reflected by the motion blur, due to occlusion in the vision-based hand gesture sensing approach. While (b) affords a form of "stable" grasp (i.e., it removes the gravity from the cup) by directly attaching the object to the hand, this approach is unnatural, with minimal realism. It does not reflect the actual contact between a hand and an object, and sometimes the hand even fails to come into contact with the object. The proposed reconfigurable glove in VR mode offers a realistic and stable grasp, which is crucial for obtaining contact points during interactions. Reproduced from Ref. [23] with permission.

while offering sufficient realism; an example is provided in Fig. 6 (c) [23].

Thanks to the reconfigurable nature of the glove, creating a virtual hand model in VR is simply the reiteration of the hand gesture-sensing module described in Section 2; Fig. 7 [23] shows the structure of the virtual hand. More specifically, the hand gestures in the local frames are given by the IMUs, and a VIVE tracker with HTC Lighthouse provides the precise positioning of the hand in a global coordinate, computed by the time-difference-of-arrival.

4.2. Stable grasps

Methods for realizing a virtual grasp in VR can be roughly categorized into two streams, with their unique pros and cons. One approach is to use a physics-based simulation with collision detection to support realistic manipulations by simulating the contact between a soft hand and a virtual object made from varied materials. Despite its high fidelity, this approach often demands a significant amount of computation, making it difficult—if not impossible—to use in real time. Alternatively, symbolic-based and rule-based grasps are popular approaches. A grasp or release is triggered based on a set of predefined rules when specific conditions are satisfied. This approach is computationally efficient but provides minimal realism.

Our configurable glove-based system must balance the above two factors to obtain contact points during interactions. It must provide a more natural interaction than those of rule-based methods, such that the contact points obtained on the objects are relatively accurate, while ensuring more effective computation than high-fidelity physics-based simulations, such that it can be achieved in real time.

In this work, we devise a caging-based stable grasp algorithm, which can be summarized as follows. First, the algorithm detects all collisions between the hands and objects (e.g., the red areas in Fig. 8(a) [23]). Next, the algorithm computes the geometric center of all collision points between the hands and objects and checks whether this center is within the object. Supposing that the above situation holds (Fig. 8(b) [23]), we consider this object to be "caged"; thus, it can be stably grasped. The objects' physical properties are turned off, allowing them to move along with the hand. Otherwise, only standard collisions are triggered between the hand and object. Finally, the grasped object is released when the collision event ends or the geometric center of the collisions is outside the object. This process ensures that a grasp only starts after a

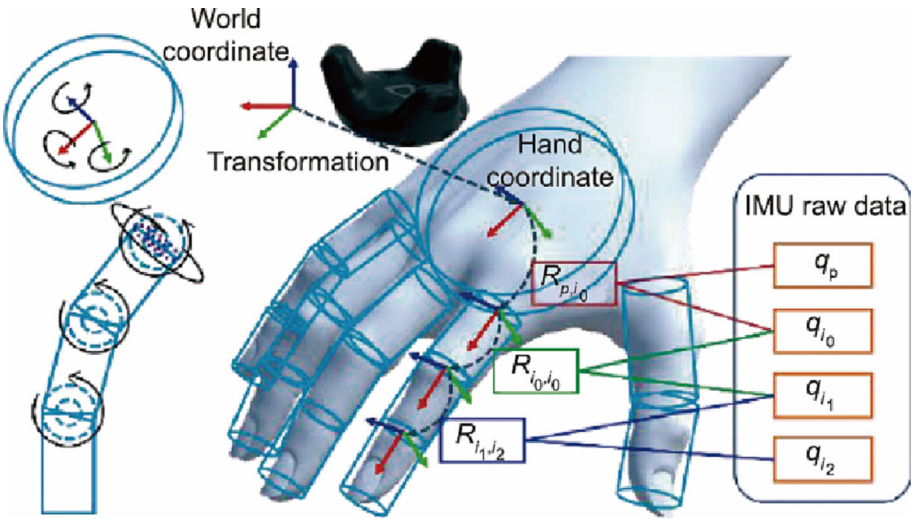


Fig. 7. Structure of the virtual hand model. Each phalanx is modeled by a small cylinder whose dimensions are measured by a participant. The pose of each phalanx is reconstructed from the data read by the IMUs. The VIVE tracker provides direct tracking of the hand pose. $q_{(.)}$: the quaternions of the IMU placed on the palm or on the proximal, middle, and distal phalanx. $R_{(.)}$: the rotation between two consecutive parts of the hand converted from their quaternions. Reproduced from Ref. [23] with permission.

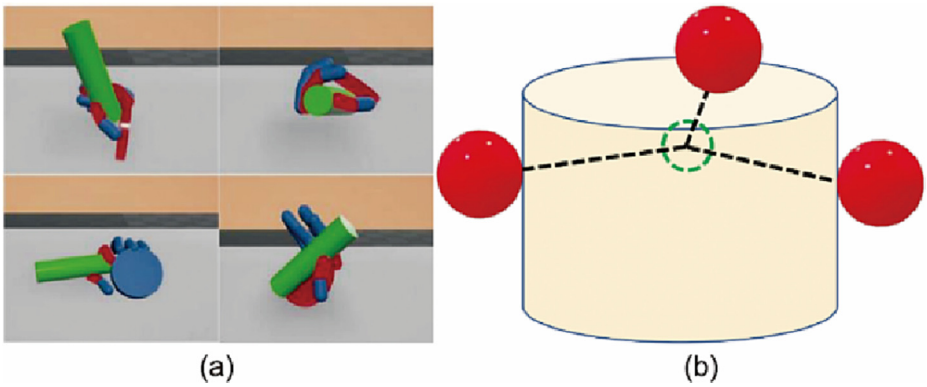


Fig. 8. Detecting a stable grasp based on collisions. (a) Various grasps of a small green cylinder. The red regions of the hand indicate the contacts with the object. (b) When the geometric center (green dashed circle) of all the collision points (red balls) overlaps with the object (yellow cylinder), the object is considered to be stably grasped and will move along with the hand. Reproduced from Ref. [23] with permission.

caging is formed, offering a more natural manipulation experience with higher realism than rule-based grasps.

4.3. Haptic feedback

By default, the participants have no way to feel whether or not their virtual hands are in contact with the virtual objects while operating the glove in VR mode due to the lack of haptic feedback, which prevents they from manipulating objects naturally. To fill this gap, the VR mode implements a network of shaftless vibration motors that are triggered when the corresponding virtual phalanxes collide with the virtual object; this offers an effective means of providing each finger with vibrational haptic feedback in the physical world that corresponds to the contact feedback that the participants should receive in VR. Connected to a 74HC4051 analog multiplexer and controlled by the RPi's GPIO, these small (10 mm × 2 mm) and lightweight (0.8 g) vibration motors provide. 14 500 r·min^{-1} with a 3 V input voltage. Once a finger touches the virtual object, the vibration motors located at that region of the glove are activated to provide continuous feedback. When the hand forms a stable grasp, all motors are powered up, so that the user can maintain the current hand gesture to hold the object.

4.4. Qualitative evaluation

We conducted a case study wherein the participants were asked to wear the VR glove and grasp four virtual objects with different shapes and functions, including a mug, a tennis racket, a bowl, and a goose toy (Fig. 9 [23]). These four objects were selected because ① they are everyday objects with a large variation in their geometry, providing a more comprehensive assessment of the virtual grasp; and ② each of the four objects can be grasped in different manners based on their functions, covering more grasp types [59,60]. We started by testing different ways of interacting with virtual objects, such as grasping a mug by either the handle or the rim. Such diverse interactions afforded a natural experience by integrating unconstrained fine-grained gestures, which is difficult for existing platforms (e.g., LeapMotion). In comparison, our reconfigurable glove in VR mode successfully balanced the naturalness of the interactions with the stability of the grasp, providing a better realism in VR, which was close to how objects are manipulated in the physical world.

Notably, the reconfigurable glove in VR mode was able to track hand gestures and maintain a stable grasp even when the hand was outside the participant's FoV, thus offering a significant advantage compared with vision-based approaches (e.g., the LeapMotion sensor). In a comparative study in which the participant's hand could be outside of the FoV, the performance using the VR glove significantly surpassed that of LeapMotion (Table 4), thereby demonstrating the efficacy of the VR glove hardware, the caging-based grasp approach, and the haptic feedback.

5. Simulation mode

A manipulation event consists of both hand information and object information. Most prior work has focused on the former without paying much attention to the latter. In fact, objects may be occluded or may even change significantly in shape as a result of a manipulation event, such as through deformation or cracking. Such information is essential in understanding the manipulation event, as it reflects the goals. However, existing solutions, even those with specialized sensors, fall short in handling this scenario, so a solution beyond the conventional scope of data gloves is called for.

To tackle this challenge, we integrate a state-of-the-art FEM simulator [19] to reconstruct the physical effects of an object, in numeric terms, during the manipulation. Given the trajectory data obtained by the proposed glove-based system, both physical and virtual properties and how they evolve over time are simulated and rendered, providing a new dimension for understanding complex manipulation events.

5.1. Simulation method

We start with a brief background of solid simulation. Solid simulation is often conducted with FEM [61], which discretizes each object into small elements with a discrete set of sample points as the DoFs. Then, mass and momentum conservation equations are discretized on the mesh and integrated over time to capture the dynamics, in which *elasticity* and *contact* are the most essential yet most challenging components. *Elasticity* is the ability of an object to retain its rest shape under external impulses or forces, whereas *contact* describes the intersection-free constraints on an object's motion trajectory. However, elasticity is nonlinear and non-convex, and contact is non-smooth, both of which can pose significant difficulties to traditional solid simulators based on numerical methods [62]. Recently, Li et al. [19] proposed incremental potential contact (IPC), a robust and accurate contact-handling method for FEM simulations [63–67]; it formulates the non-smooth contact condition into smooth approximate barrier potentials so that the non-smooth contact condition can be solved simultaneously with electrodynamics using a line search method [68–70] with a global convergence guarantee. As it is able to

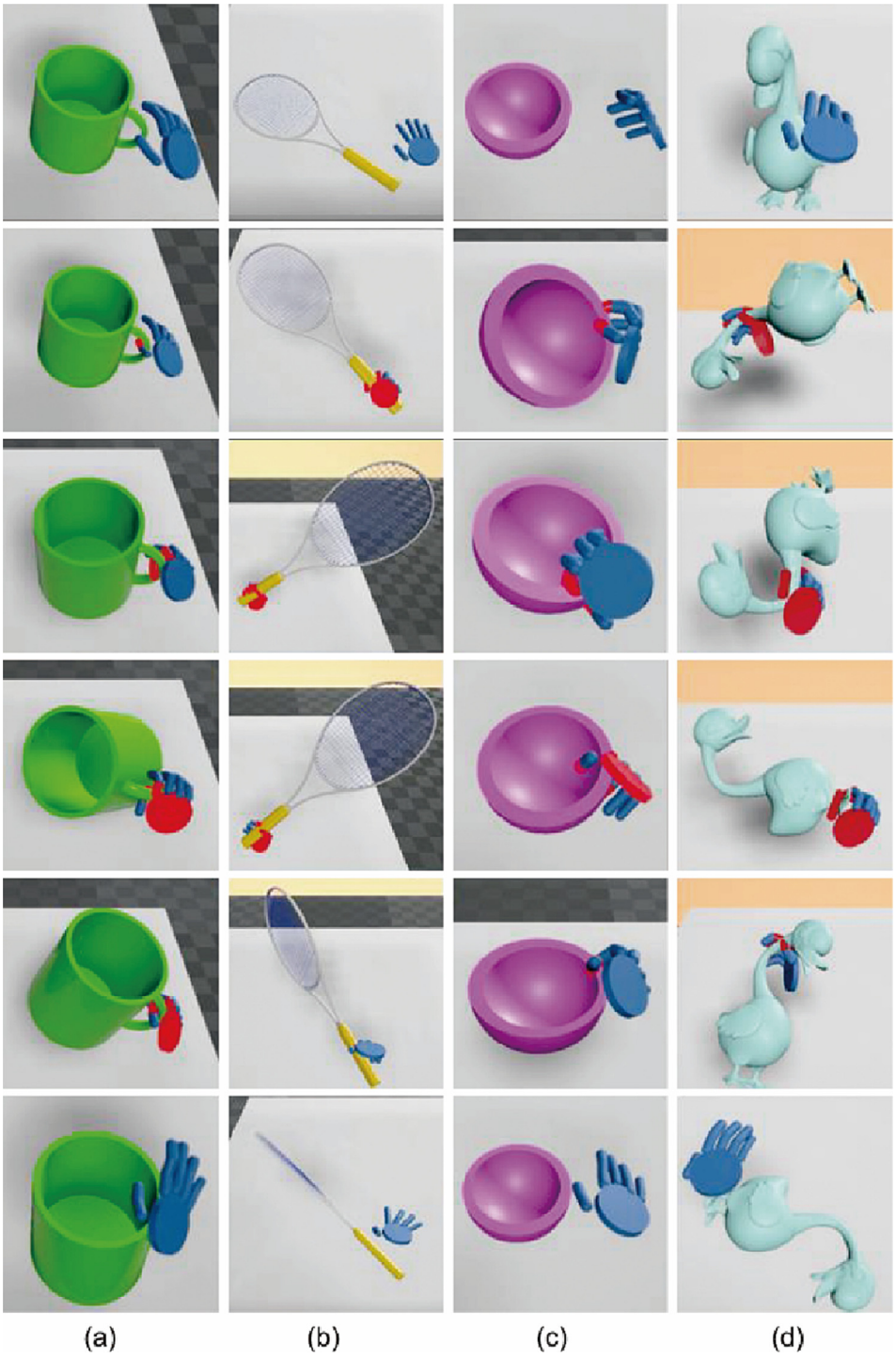


Fig. 9. Various grasp results for four virtual objects: (a) a mug, (b) a tennis racket, (c) a bowl, and (d) a goose toy. The top and bottom rows show the approach and release of the target objects, respectively. Reproduced from Ref. [23] with permission.

Table 4
Success rates of grasping and moving four different objects using the VR glove and the LeapMotion sensor.

Task	Setup	Mug	Racket	Mug	Racket
Grasp	LeapMotion sensor	80%	13%	27%	67%
	VR glove	100%	100%	100%	93%
Move	LeapMotion sensor	33%	7%	0	47%
	VR glove	100%	93%	93%	87%

consistently produce high-quality results without numerical instability issues, IPC makes it possible to conveniently simulate complex manipulation events, even with extremely large deformations.

We further extend the original IPC to support object fracture by measuring the displacement of every pair of points; that is, we go through all pairs of points for a triangle and all triangles on the mesh. If the displacement relative to the pair of points' original distance exceeds a certain strain threshold (in this work, we set it to 1.1), we mark the triangle in between as separated. At the end of every time step, we reconstruct the mesh topology using a graph-based approach [71], according to the tetrahedra face separation information. Due to the existence of the IPC barrier, which only allows a positive distance between surface primitives, it is essential to ensure that, after the topology change, the split faces do not exactly overlap. Therefore, we perturb the duplicate nodes on the split faces by a tiny displacement toward the normal direction, which works nicely even when edge-edge contact pairs are ignored for simplicity.

5.2. Prototyping and input data collection

The simulation-augmented glove-based system is essentially the same as the VR glove, except for the lack of vibration motors; however, it is augmented with the simulated force evolved over time. Compared with the aforementioned two hardware-focused designs, the simulation-augmented glove-based system offers an in-depth prediction of physics with fine-grained object dynamics—that is, how the geometry (e.g., large deformation) and topology (e.g., fracture) evolve. To showcase the efficacy of this system, we focus on a tool-use setting wherein a user manipulates a tool (e.g., a hammer) to apply on a target object (e.g., a nut), causing geometry and/or topology changes. To collect one set of data, the hand gestures and poses are reconstructed similarly using the other two glove-base systems. The tool's movement is further tracked to simulate the interactions between the tool and the object.

More specifically, two VIVE trackers track the movements of the glove-based system (i.e., the hand) and the tool, respectively. The third tracker, which serves as the reference point for the target object (e.g., a nut) is fixed to the table. All three VIVE trackers are calibrated such that their relative poses and the captured trajectories can be expressed in the same coordinate. The target objects and the tool's meshes are scanned beforehand using a depth camera. By combining the scanned meshes and captured trajectories, we can fully reconstruct a sequence of 3D meshes representing the movements of the hand and tool and simulate the resulting physical effects of the target object. The captured mesh sequences are directly input to the simulation as boundary conditions, and the DoFs being simulated are primarily those on the target object. Fig. 10 shows some keyframes of the data collection for cracking walnuts and cutting carrots. It should be noted that capturing how the object changes and its physical properties over time is extremely challenging—if not impossible—using visual information alone.

5.3. Simulation setup

An object's material properties in a simulation are mainly reflected by its stiffness (i.e., the object is more difficult to deform or fracture if it is stiffer), governed by its Young's modulus and Poisson's ratio. These parameters must be set appropriately in the simulation in order to produce effects that match those in the physical world. The Young's modulus and Poisson's ratio of a material can be found in related works [72–74]. Another parameter that must be set is the fracturing strain threshold, which determines the dimension of the segments when fracturing is triggered. This parameter is tuned so that the simulator can reproduce the type of effects observed in the physical world. The time step of the simulation is the inversion of the sampling frequency of the Vive trackers that acquire the trajectories.

6. Application

In this section, we showcase a series of applications by reconfiguring the data glove to the tactile-sensing mode (Section 6.1), VR mode (Section 6.2), and simulation mode (Section 6.3), all of which share the same backbone design (video demonstrations in the Appendix A).

6.1. Tactile-sensing mode

We evaluated the tactile-sensing mode by capturing the manipulation data of opening three types of medicine bottles. Two of these bottles are equipped with different locking mechanisms and require a series of specific action sequences to remove the lid. More specifically, Bottle 1 does not have a safety lock, and simply twisting the lid is sufficient to open it. The lid of Bottle 2 must be pressed simultaneously while twisting it. Bottle 3 has a safety lock in its lid, which requires a pinching action before twisting to unlock it. Notably, the pressing and pinching actions required to open Bottle 2 and Bottle 3 are challenging to recognize without using the force information recorded by the glove.

Fig. 11 [20] shows examples of the recorded data with both hand gesture and force information. The first row of Fig. 11 [20]

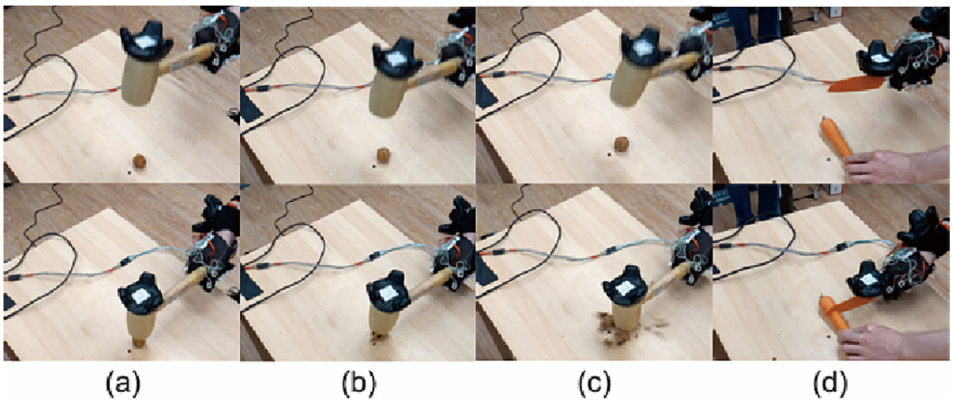


Fig. 10. Four types of tool-use events captured by a slow-motion camera at 120 frames per second (fps). These categories, in terms of fluent changes, are: (a) uncracked, (b) cracked, (c) smashed, and (d) cut in half.

visualizes the captured manipulation action sequences of opening these three bottles. The second row shows the corresponding action sequences captured by a red–green–blue (RGB) camera for reference.

Qualitatively, compared with the action sequences shown in the second row, the visualization results in the first row differentiate the fine manipulation actions with additional force information. For example, the fingers in Fig. 11(b) [20] are flat and parallel to the bottle lid, whereas those in Fig. 11(c) [20] are similar to those in the gripping pose. The responses of the force markers are also different due to varying contact points between the human hand and the lid: The high responses in Fig. 11(b) are concentrated on the palm area, whereas only two evident responses on the distal thumb and index finger can be seen in Fig. 11(c) [20]. Taken together, these results demonstrate the significance of accounting for forces when understanding fine manipulation actions.

Quantitatively, Fig. 12 [20] illustrates one taxel's force collected on the palm, the thumb's fingertip, and the flexion angle of the index finger's MCP joint. In combination, these three readings can differentiate among the action sequences of opening the three bottles. More specifically, as opening Bottle 2 involves a pressing action on the lid, the tactile glove successfully captures the high force response on the palm. In contrast, the force reading in the same region is almost zero when opening the other two bottles. Bottle 3′s pinch-to-open lock necessitates a greater force exerted by the thumb. Indeed, the opening actions introduce a high force response at the thumb's fingertip, with a longer duration than the actions involved in opening Bottle 1 without a safety lock. Without contacting the lid, the thumb yields no force response when opening Bottle 2. Since opening both Bottle 1 and Bottle 3 involves a similar twist action, the measured flexion angles of the index finger's MCP joint are around 50° in both of these cases.

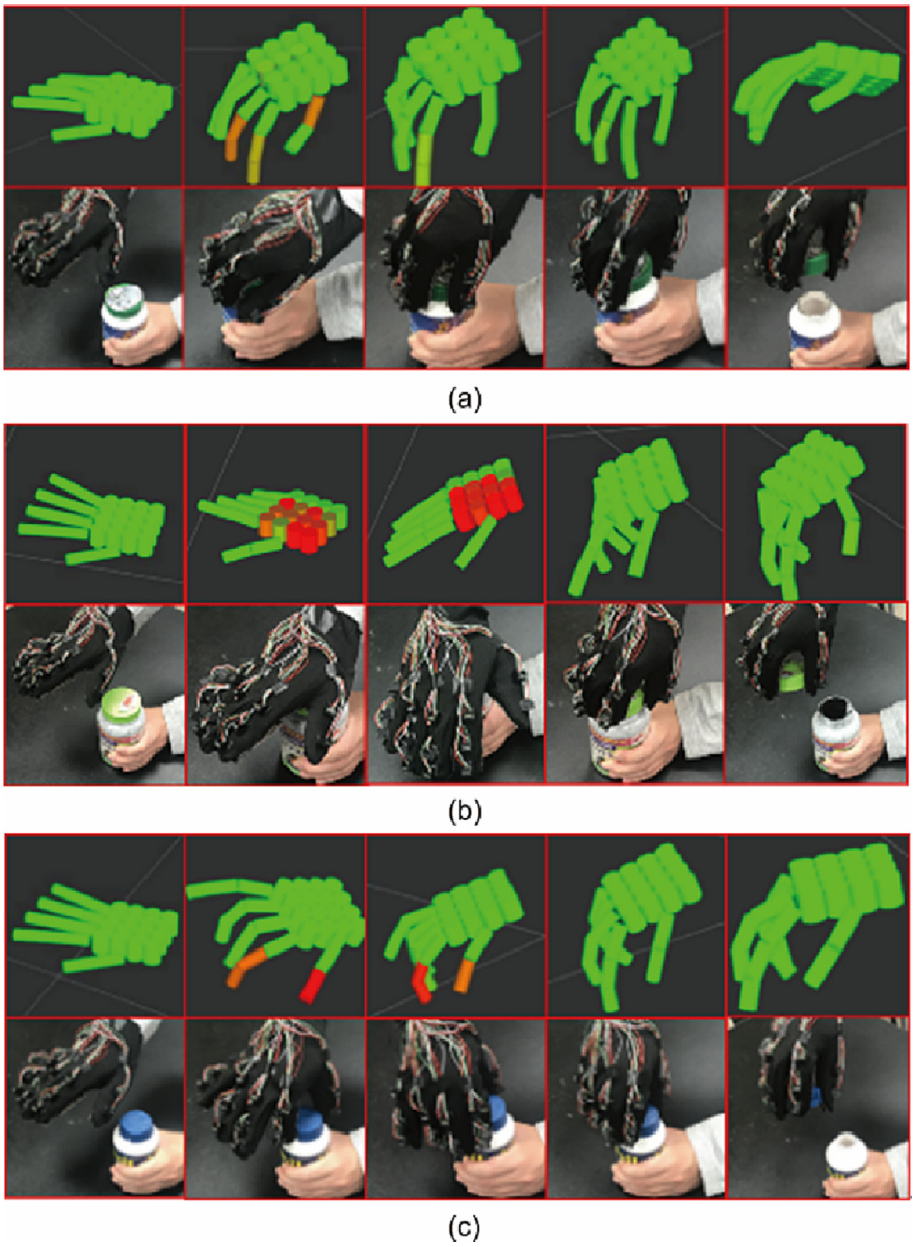


Fig. 11. Visualizations of the hand gesture and force of opening three bottles, (a) Bottle 1, no childproof lock; (b) Bottle 2, pressing down the lid to unlock; (c) Bottle 3, pinching the lid to unlock, collected using the tactile-sensing glove. These visualizations reveal the subtle differences between the actions of opening medicine bottles and opening a conventional bottle; the essence of this task is that visual information alone is insufficient to distinguish between the opening of the various bottles. Reproduced from Ref. [20] with permission.

Since only the palm touches the lid and the fingers remain stretched, a small flexion angle occurs when opening Bottle 2.

A promising application of the proposed glove is learning fine manipulation actions from human demonstrations. The collected tactile data has facilitated investigations into a robot's functional understanding of actions and imitation learning [12,75], inverse reinforcement learning [76], and learning explainable models that promote human trust [21]. Fig. 13 [15] showcases the robot's learned skills of opening different medicine bottles [75].

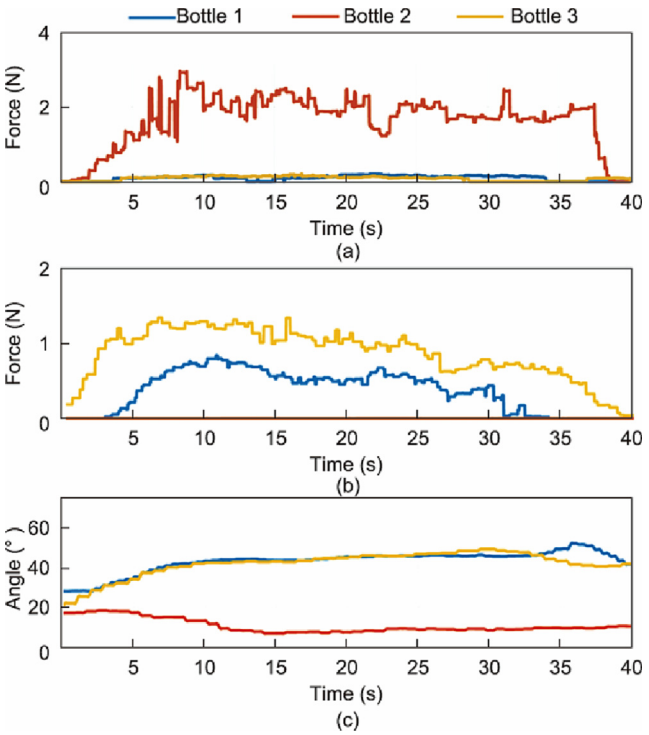


Fig. 12. Force and joint angle recorded by the tactile-sensing glove. (a) The forces exerted by the palm, (b) forces exerted by the thumb's fingertip, and (c) the flexion angle of the index finger's MCP joint can disentangle the grasp actions of opening different bottles. Reproduced from Ref. [20] with permission.

6.2. VR mode

When operating in VR mode, the reconfigurable glove provides a unique advantage compared with traditional hardware. Below, we showcase two data types that can be collected effectively in this mode.

6.2.1. Trajectories

Hand and object trajectories are particularly useful in robot learning from demonstration. Diverse object models can be placed in the VR without setting up a physical apparatus to ensure a natural hand trajectory. Fig. 14 [23] shows some qualitative results of collected trajectories: the hand movement (red line) and the five fingertips' trajectories (blue lines) by combining global hand pose and hand gesture sensing, and the grasped object's movement (black line) as the result of hand movement and grasp configuration (stable grasp or not). These results demonstrate the reliability of our design and the richness of the collected trajectory information in a manipulation event.

6.2.2. Contact points

It is extremely challenging to obtain the contact points of the objects being manipulated. Despite relying heavily on training data, computer vision-based methods [77] are still vulnerable to handling occlusion between hands and objects. Our reconfigurable glove operating in the VR mode can elegantly log this type of data. Given the meshes of the virtual hand model and the object, the VR's physics engine can effectively check the collisions between them. These collisions not only determine whether the object can be stably grasped based on the criteria described in Section 4.2 but also correspond well to the contact points on the grasped object. By treating a collision point as the spatial center of a spherical volume whose radius is set to the diameter of the finger, Fig. 15 shows three configurations of contacts collected from different participants grasping diverse objects. To better uncover the general grasp habits for an object, the contact points shown in the bottom row of Fig. 15 are obtained by averaging the spatial positions of contacts across different trails, fitted by a Gaussian distribution.

A fundamental challenge in robot learning of manipulation is the embodiment problem [12,78]: The human hand (five fingers) and robot gripper (usually two or three fingers) have different morphologies. While this problem demands further research, individ-

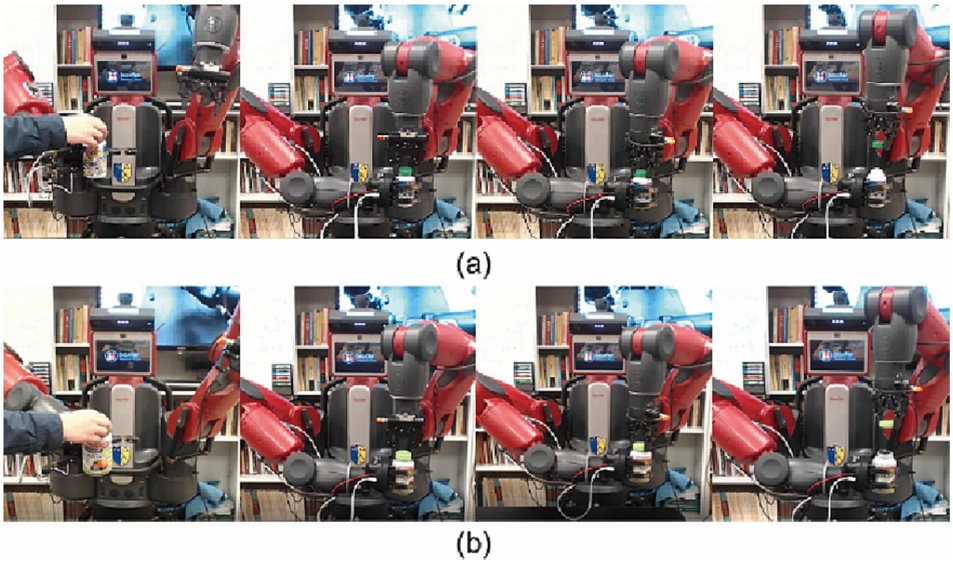


Fig. 13. A Baxter robot learns to open medicine bottles from the collected manipulation data. Reproduced from Ref. [75] with permission.

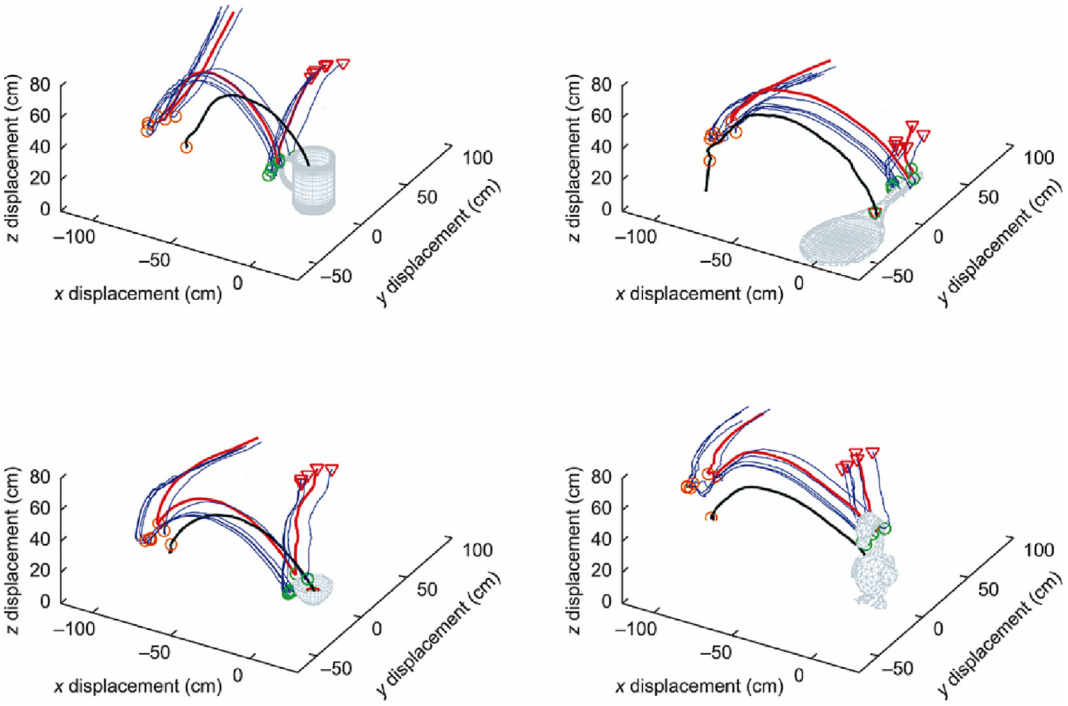


Fig. 14. Examples of hand and object trajectories collected by the reconfigurable glove operating in VR mode. Red triangles indicate the starting poses. The red line and the blue lines show the recorded hand movement and the trajectories of the fingertips, respectively. Once the contact points (green circles) are sufficient to trigger a stable grasp, the object moves together with the hand, following the black line, until the grasp becomes unstable—that is, until it is released at the orange circles. Reproduced from Ref. [23] with permission.

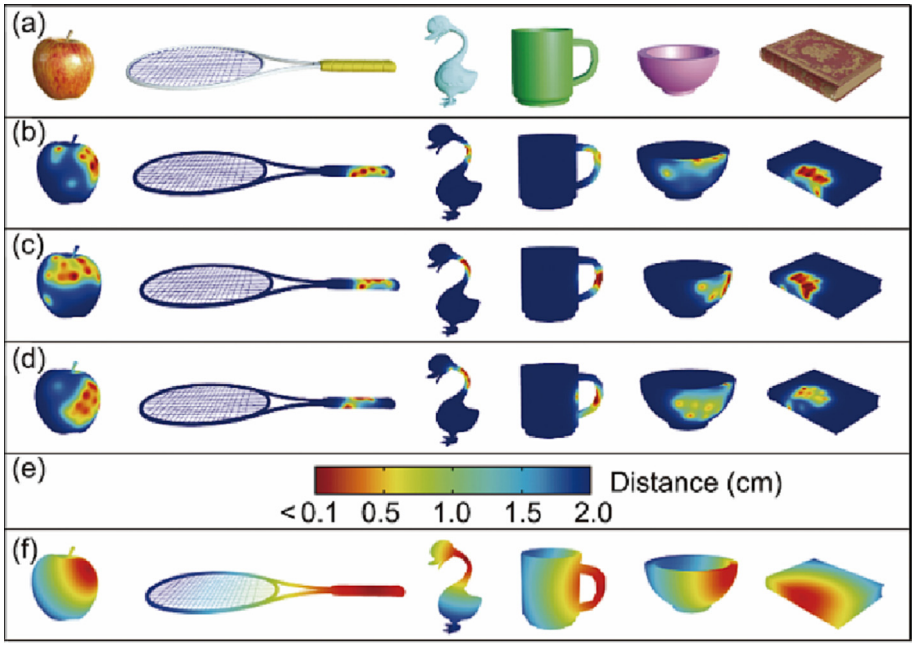


Fig. 15. Contact points in grasping various objects. (a) Objects to be grasped. (b–d) Three configurations of the contact points performed by different participants. (e) The distance from each contact point. (f) The average of contact points aggregated from all participants, indicating the preferred regions of contact, given the objects.

ual contact points can also indicate a preferred region of contact if aggregated from different participants (see the last row in Fig. 15). Such aggregated data can be used for training robot manipulation policies despite different morphologies [12].

6.3. Simulation mode

By incorporating the state-of-the-art physics-based simulation, we empower the data glove to capture fine-grained object dynam-

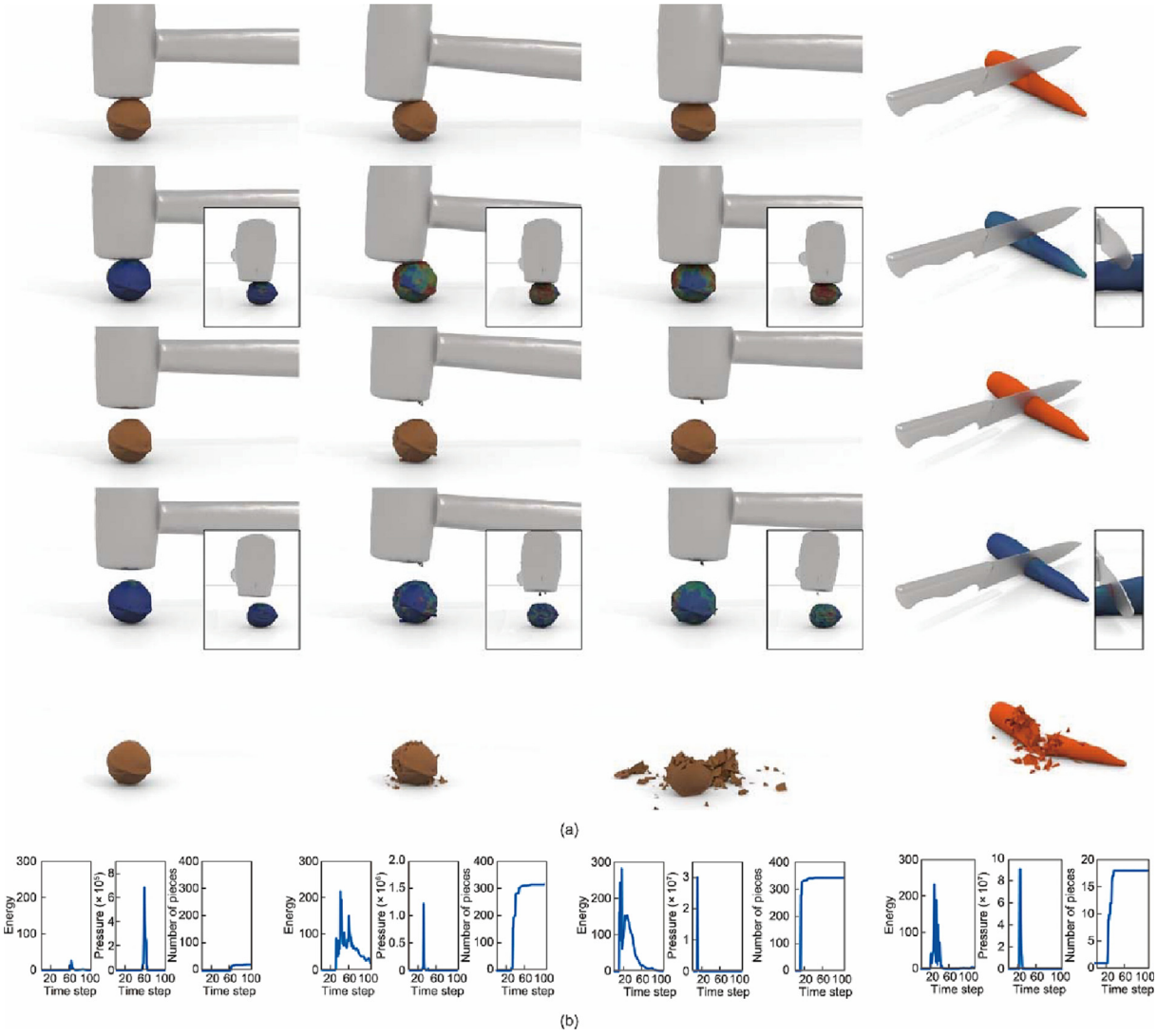


Fig. 16. Reconstructed 4D manipulation events of tool use by integrating trajectories collected by the reconfigurable glove and physics-based simulation. This high-fidelity 4D data reveals fine-grained object fluent changes and physical properties at each time step. The results are produced with a simulation at 20 Hz; one time step is 0.05 s. (a) Reconstructed tool-use events by simulation. The first/third rows show the contact moments between the tool and the object. The second/fourth rows are the corresponding stress given by the simulator; red indicates greater stress. The fifth row shows the objects' final status. (b) The energy imposed on the objects, the number of fractured pieces, and the contact pressure calculated by the simulator at each time step during tool use.

ics during manipulations. Fig. 16 showcases simulated objects' fluent changes in tool uses. Even when recorded at 120 fps, it is challenging—if not impossible—to capture an object's fluent changes (e.g., how a walnut smashes) using a vision-based method. By feeding the collected trajectory into the simulation, our system renders object fluent changes that are visually similar to the physical reality (Fig. 16(a)), thereby revealing critical physical information (Fig. 16(b)) on what occurs in the process.

6.3.1. Results

Fig. 16(a) depicts various processes of hammering a walnut. The first column illustrates that a gentle swing action only introduces a small force/energy to the walnut, resulting in a light stress distribution that is quickly eliminated; as a result, the walnut remains uncracked. When a strong swing is performed (third column in Fig. 16(a)), the larger internal stress causes the walnut to fracture into many pieces, similar to a smashing event in the physical world. This difference is reflected in Fig. 16(b), which was obtained using the physics-based simulator. It is notable that these physical quantities are challenging to measure in the physical world, even with specialized equipment.

6.3.2. Failure examples

The fourth column of Fig. 16(a) shows an example of cutting a carrot. The imposed stress is concentrated along the blade that splits the carrot in half. However, when the cutting action is completed and the knife is lifted, it can be seen that the collision between the blade and the carrot has caused undesired fracturing around the cut, which illustrates the limit of the current simulator.

7. Discussion

We now discuss two topics in greater depth: Are simulated results good enough, and how do the simulated results help?

7.1. Are simulated results good enough?

A central question regarding simulations is whether the simulated results are helpful, given that they are not numerically iden-

tical to those directly measured in the physical world. We argue that simulators are indeed helpful, as a simulation preserves the physical events qualitatively, making it possible to study complex events. As illustrated in Fig. 16(b), the walnut's effects have a clear correspondence to the pressure imposed on the contact. Conversely, although a similar amount of energy is imposed when cracking the walnut with a hammer and cutting the carrot with a knife (see the second and fourth columns of Fig. 16), the resulting pressures differ in magnitude, as the knife introduces a much smaller contact area than the hammer does, producing distinct deformations and topology changes. Hence, the simulation provides a qualitative measurement of the physical events and the objects' fluent change rather than precise quantities. Similar arguments are found in the *intuitive physics* literature in psychology: Humans usually only make approximate predictions about how states evolve, sometimes even with violations of actual physical laws [79]. Such inaccuracy does not prevent humans from possessing an effective object and scene understanding; on the contrary, it is a core component of human commonsense knowledge [80–82]. Recent work in robot tool use [83–85] and physics-informed scene understanding [86–94] has also demonstrated the essential role of physics in understanding objects and scenes.

7.2. How do the simulated results help?

The fine-grained object effects produced by the simulation open up new venues for studying existing AI and robotics problems. For example, combining task planning and motion planning [95–97] is a grand challenge in the field of planning. Simulation could help with this challenge in two aspects [83]: ① by grounding ambiguous task symbols to desired outcomes (e.g., the action symbol of "crack"), and ② by modeling implicit goal specifications (e.g., the status of "cracked"). In addition, simulations can be used to augment existing datasets, such as GARB [98] and GenDexGrasp [84] in grasping and HUMANISE [99], CHAIRS [100], and LEMMA [101] in scene understanding with unobservable information. Ultimately, we hope that this type of 4D data empowered by physics-based simulation can shed light on several profound questions in manipulation: What and why an object is chosen (i.e., the physics involved), how to properly operate that object (i.e., its affordance), what effect the actor is trying to achieve (i.e., the actor's task goals), and what happens when the goal is not achieved (i.e., planning and replanning).

8. Conclusions

In this study, we presented three different configurations of a glove-based system based on a unified backbone design, which differs from most conventional data gloves that only capture hand gestures. Utilizing piezoresistive Velostat material, the glove's tactile-sensing mode can aggregate the hand force information during manipulation events. In VR mode, the sensed hand gestures can be reconstructed into a virtual hand to facilitate hand–object interactions in VR by incorporating a caging-based approach, resulting in stable grasps and providing vibrational haptic feedback. The simulation mode further uses an FEM simulator to produce fine-grained object fluent changes and physical properties based on hand-related movements, resulting in 4D manipulation events.

We evaluated the components of the system, including the IMUs, Velostat force-sensor taxels, and haptic feedback provided by the vibration motors, to demonstrate the capability and efficacy of the proposed design. By ① capturing spatiotemporal signals of force and gesture, ② recording hand trajectories and contact points on objects, and ③ collecting 4D manipulations in challenging manipulation events (e.g., tool use), we demonstrated that the proposed glove-based system can play a crucial role in robot learning from humans and in facilitating embodied AI-related research.

Acknowledgments

The authors would like to thank Mr. Matt Millar and Dr. Xu Xie (Meta) for developing earlier versions of the system, Miss Chen Zhen (BIGAI) for making the nice figures, and five anonymous reviews for constructive feedback. This work is supported in part by the National Key R&D Program of China (2021ZD0150200) and the Beijing Nova Program.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.eng.2023.01.009.

References

[1] Pinto L, Gupta A. Supersizing self-supervision: learning to grasp from 50K tries and 700 robot hours. In: Proceedings of IEEE International Conference on Robotics and Automation (ICRA 2016); 2016 May 16–21; Stockholm, Sweden. New York City: IEEE; 2016.
[2] Mahler J, Matl M, Satish V, Danielczuk M, DeRose B, McKinley S, et al. Learning ambidextrous robot grasping policies. Sci Robot 2019;4(26):eaau4984.
[3] Zeng A, Song S, Yu KT, Donlon E, Hogan FR, Bauza M, et al. Robotic pick-and-place of novel objects in clutter with multi-affordance grasping and cross-domain image matching. In: Proceedings of IEEE International Conference on Robotics and Automation (ICRA 2018); 2018 May 21–25; Brisbane, QLD, Australia. New York City: IEEE; 2018.
[4] Cini F, Ortenzi V, Corke P, Controzzi M. On the choice of grasp type and location when handing over an object. Sci Robot 2019;4(27):eaau9757.
[5] Yahya A, Li A, Kalakrishnan M, Chebotar Y, Levine S. Collective robot reinforcement learning with distributed asynchronous guided policy search. In: Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2017); 2017 Sep 24–28; Vancouver, BC, Canada. New York City: IEEE; 2017. p. 79–86.
[6] Schaal S, Ijspeert A, Billard A. Computational approaches to motor learning by imitation. Phil Trans R Soc Lond B 2003;358(1431):537–47.
[7] Maeda G, Ewerton M, Koert D, Peters J. Acquiring and generalizing the embodiment mapping from human observations to robot skills. IEEE Robot Autom Lett 2016;1(2):784–91.
[8] Nguyen A, Kanoulas D, Caldwell DG, Tsagarakis NG. Detecting object affordances with convolutional neural networks. In: Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2016); 2016 Oct 9–14; Daejeon, Republic of Korea. New York City: IEEE; 2016. p. 2765–70.
[9] Kokic M. Stork JA, Haustein JA, Kragic D. Affordance detection for task-specific grasping using deep learning. In: 2017 IEEE-RAS 17th International Conference on Humanoid Robotics (Humanoids); 2017 Nov 15–17; Birmingham, UK. New York City: IEEE; 2017. p. 91–8.
[10] Mohseni-Kabir A, Rich C, Chernova S, Sidner CL, Miller D. Interactive hierarchical task learning from a single demonstration. In: Proceedings of the 2015 10th Annual ACM/IEEE International Conference on Human–Robot Interaction; 2015 Mar 2–5; Portland, OR, USA. New York City: IEEE; 2015. p. 205–12.
[11] Xiong C, Shukla N, Xiong W, Zhu SC. Robot learning with a spatial, temporal, and causal and–or graph. In: Proceedings of 2016 IEEE International Conference on Robotics and Automation (ICRA 2016); 2016 May 16–21; Stockholm, Sweden. New York City: IEEE; 2016. p. 2144–51.
[12] Liu H, Zhang C, Zhu Y, Jiang C, Zhu SC. Mirroring without overimitation: learning functionally equivalent manipulation actions. In: Proceedings of the AAAI Conference on Artificial Intelligence (AAAI); 2019 Jan 27–Feb 1; Honolulu, HI, USA. 2019. p. 8025-8033.
[13] Abbeel P, Ng AY. Apprenticeship learning via inverse reinforcement learning. In: Proceedings of the 21st International Conference on Machine Learning (ICML 2004); 2004 Jul 4–8; Banff, AB, Canada. New York City: Association for Computing Machinery (ACM); 2004.
[14] Prieur U, Perdereau V, Bernardino A. Modeling and planning high-level in-hand manipulation actions from human knowledge and active learning from demonstration. In: Proceedings of 2012 IEEE/RSJ International Conference on Intelligent Robots and Systems; 2012 Oct 7–12; Vilamoura-Algarve, Portugal. New York City: IEEE; 2012. p. 1330–6.
[15] Ibarz B, Leike J, Pohlen T, Irving G, Legg S, Amodei D. Reward learning from human preferences and demonstrations in Atari. In: Proceedings of the 32nd Conference on Advances in Neural Information Processing Systems (NeurIPS 2018); 2018 Dec 3–8; Montréal, QC, Canada. Red Hook: Curran Associates Inc.; 2018. p. 1–13.

[16] Xie X, Liu H, Zhang Z, Qiu Y, Gao F, Qi S, et al. VRGym: a virtual testbed for physical and interactive AI. In: Proceedings of the ACM Turing Celebration Conference-China; 2019 May 17–19; Chengdu, China. New York City: Association for Computing Machinery; 2019. p. 1–6.
[17] Li C, Xia F, Martín-Martín R, Lingelbach M, Srivastava S, Shen B, et al. IGibson 2.0: object-centric simulation for robot learning of everyday household tasks. In: Proceedings of the 5th Annual Conference on Robot Learning (CoRL 2021); 2021 Nov 8; London, UK; online; 2021.
[18] Szot A, Clegg A, Undersander E, Wijmans E, Zhao Y, Turner J, et al. Habitat 2.0: training home assistants to rearrange their habitat. In: Proceedings of 35th Conference on Neural Information Processing Systems (NeurIPS 2021); 2021 Dec 6–14; online; 2021.
[19] Li M, Ferguson Z, Schneider T, Langlois T, Zorin D, Panozzo D, et al. Incremental potential contact: intersection-and inversion-free, large-deformation dynamics. ACM Trans Graph 2020;39(4):49.
[20] Liu H, Xie X, Millar M, Edmonds M, Gao F, Zhu Y, et al. A glove-based system for studying hand–object manipulation via joint pose and force sensing. In: Proceedings of 2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS); 2019 Sep 24–28; Vancouver, BC, Canada. New York City: IEEE; 2017. p. 6617–24.
[21] Edmonds M, Gao F, Liu H, Xie X, Qi S, Rothrock B, et al. A tale of two explanations: enhancing human trust by explaining robot behavior. Sci Robot 2019;4(37):aay4663.
[22] Brahmbhatt S, Ham C, Kemp CC, Hays J. ContactDB: analyzing and predicting grasp contact via thermal imaging. In: Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR 2019); 2019 Jun 15–20; Long Beach, CA, USA. New York City: IEEE; 2019. p. 8701–11.
[23] Liu H, Zhang Z, Xie X, Zhu Y, Liu Y, Wang Y, et al. High-fidelity grasping in virtual reality using a glove-based system. In: Proceedings of the 2019 International Conference on Robotics and Automation (ICRA 2019); 2019 May 20–24; Montreal, QC, Canada. New York City: IEEE; 2019. p. 5180–6.
[24] Duan K, Parikh D, Crandall D, Grauman K. Discovering localized attributes for fine-grained recognition. In: Proceedings of the 2012 IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2012); 2012 Jun 16–21; Providence, RI, USA. New York City: IEEE; 2012. p. 3474–81.
[25] Liu Y, Wei P, Zhu SC. Jointly recognizing object fluents and tasks in egocentric videos. In: Proceedings of 2017 IEEE International Conference on Computer Vision (ICCV); 2017 Oct 22–29; Venice, Italy. New York City: IEEE; 2017. p. 2943–51.
[26] Nagarajan T, Grauman K. Attributes as operators: factorizing unseen attribute-object compositions. In: Proceedings of European Conference on Computer Vision (ECCV 2018); 2018 Sep 8–14; Munich, Germany. Berlin: Springer; 2018. p. 172–90.
[27] Newton I, Colson J. The method of fluxions and infinite series; with its application to the geometry of curve-lines. London: Henry Woodfall; 1736..
[28] Dipietro L, Sabatini AM, Dario P. A survey of glove-based systems and their applications. IEEE Trans Syst Man Cybern Part C 2008;38(4):461–82.
[29] Kramer RK, Majidi C, Sahai R, Wood RJ. Soft curvature sensors for joint angle proprioception. In: Proceedings of 2011 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2011); 2011 Sep 25–30; San Francisco, CA, USA. New York City: IEEE; 2011. p. 1919–26.
[30] Kamel NS, Sayeed S, Ellis GA. Glove-based approach to online signature verification. IEEE Trans Pattern Anal Mach Intell 2008;30(6):1109–13.
[31] Oh J, Kim S, Lee S, Jeong S, Ko SH, Bae J. A liquid metal based multimodal sensor and haptic feedback device for thermal and tactile sensation generation in virtual reality. Adv Funct Mater 2021;31(39):2007772.
[32] Wang M, Yan Z, Wang T, Cai P, Gao S, Zeng Y, et al. Gesture recognition using a bioinspired learning architecture that integrates visual data with somatosensory data from stretchable sensors. Nat Electron 2020;3(9):563–70.
[33] Wen F, Sun Z, He T, Shi Q, Zhu M, Zhang Z, et al. Machine learning glove using self-powered conductive superhydrophobic triboelectric textile for gesture recognition in VR/AR applications. Adv Sci 2020;7(14):2000261.
[34] Taylor T, Ko S, Mastrangelo C, Bamberg SJM. Forward kinematics using IMU on-body sensor network for mobile analysis of human kinematics. In: 2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC 2013); 2013 Jul 3–7; Osaka, Japan. New York City: IEEE; 2013. p. 1230–3.
[35] Kortier HG, Sluiter VI, Roetenberg D, Veltink PH. Assessment of hand kinematics using inertial and magnetic sensors. J NeuroEng Rehabil 2014;11(1):70.
[36] Hu B, Ding T, Peng Y, Liu L, Wen X. Flexible and attachable inertial measurement unit (IMU)-based motion capture instrumentation for the characterization of hand kinematics: a pilot study. Instrum Sci Technol 2020;49(2):125–45.
[37] Santaera G, Luberto E, Serio A, Gabiccini M, Bicchi A. Low-cost, fast and accurate reconstruction of robotic and human postures via IMU measurements. In: Proceedings of 2015 IEEE International Conference on Robotics and Automation (ICRA 2015); 2015 May 26–30; Seattle, WA, USA. New York City: IEEE; 2015. p. 2728–35.
[38] Ligorio G, Sabatini AM. Extended Kalman filter-based methods for pose estimation using visual, inertial and magnetic sensors: comparative analysis and performance evaluation. Sensors 2013;13(2):1919–41.
[39] Kortier HG, Antonsson J, Schepers HM, Gustafsson F, Veltink PH. Hand pose estimation by fusion of inertial and magnetic sensing aided by a permanent magnet. IEEE Trans Neural Syst Rehabiln Eng 2015;23(5):796–806.
[40] Hammond FL, Menguč Y, Wood RJ. Toward a modular soft sensor-embedded glove for human hand motion and tactile pressure measurement. In: Proceedings of 2014 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2014); 2014 Sep 14–18; Chicago, IL, USA. New York City: IEEE; 2014. p. 4000–7.
[41] Gu Y, Sheng W, Liu M, Ou Y. Fine manipulative action recognition through sensor fusion. In: Proceedings of 2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2015); 2015 Sep 28–Oct 2; Hamburg, Germany. New York City: IEEE; 2015. p. 886–91.
[42] Mohammadi M, Baldi TL, Scheggi S, Prattichizzo D. Fingertip force estimation via inertial and magnetic sensors in deformable object manipulation. In: Proceedings of the International Symposium on Haptic Interfaces for Virtual Environment and Teleoperator Systems (HAPTICS 2016); 2016 Apr 8–11; Philadelphia, PA, USA. New York City: IEEE; 2016. p. 284–9.
[43] Lin BS, Lee IJ, Chen JL. Novel assembled sensorized glove platform for comprehensive hand function assessment by using inertial sensors and force sensing resistors. IEEE Sensors J 2020;20(6):3379–89.
[44] Battaglia E, Bianchi M, Altobelli A, Grioli G, Catalano MG, Serio A, et al. ThimbleSense: a fingertip-wearable tactile sensor for grasp analysis. IEEE Trans Haptics 2016;9(1):121–33.
[45] Low JH, Khin PM, Yeow CH. A pressure-redistributing insole using soft sensors and actuators. In: Proceedings of 2015 IEEE International Conference on Robotics and Automation (ICRA 2015); 2015 May 26–30; Seattle, WA, USA. New York City: IEEE; 2015. p. 2926–30.
[46] Pugach G, Melnyk A, Tolochko O, Pitti A, Gaussier P. Touch-based admittance control of a robotic arm using neural learning of an artificial skin. In: Proceedings of 2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2016); 2016 Oct 9–14; Daejeon, Republic of Korea. New York City: IEEE; 2016. p. 3374–80.
[47] Müller S, Schröter C, Gross HM. Smart fur tactile sensor for a socially assistive mobile robot. In: Proceedings of International Conference on Intelligent Robotics and Applications (ICIRA 2015); 2015 Aug 24–27; Portsmouth, UK. Berlin: Springer; 2015. p. 49–60.
[48] Jeong E, Lee J, Kim D. Finger-gesture recognition glove using Velostat. In: Proceedings of 2011 11th International Conference on Control, Automation and Systems (ICCAS 2011); 2011 Oct 26–29; Gyeonggi-do, Republic of Korea. New York City: IEEE; 2011. p. 206–10.
[49] Boulic R, Rezzonico S, Thalmann D. Multi-finger manipulation of virtual objects. In: Proceedings of the ACM Symposium on Virtual Reality Software and Technology (VRST 1996); 1996 Jul 1–4; Hong Kong, China. New York City: Association for Computing Machinery (ACM); 1996. p. 67–74.
[50] Choi H, Crump C, Duriez C, Elmquist A, Hager G, Han D, et al. On the use of simulation in robotics: opportunities, challenges, and suggestions for moving forward. Proc Nat Acad Sci USA 2019;118(1):e1907856118.
[51] Hu Y, Liu J, Spielberg A, Tenenbaum JB, Freeman WT, Wu J, et al. ChainQueen: a real-time differentiable physical simulator for soft robotics. In: Proceedings of 2019 International Conference on Robotics and Automation (ICRA 2019); 2019 Dec 4–6; Montréal, QC, Canada. 2019. p. 6265–71.
[52] Kennedy M, Schmeckpeper K, Thakur D, Jiang C, Kumar V, Daniilidis K. Autonomous precision pouring from unknown containers. IEEE Robot Autom Lett 2019;4(3):2317–24.
[53] Heiden E, Macklin M, Narang Y, Fox D, Garg A, Ramos F. DiSECt: a differentiable simulation engine for autonomous robotic cutting. In: Proceedings of the 2021 Robotics: Science and Systems (RSS 2021); 2021 Jul 12–16; online; New York City: IEEE; 2021.
[54] Wolper J, Fang Y, Li M, Lu J, Gao M, Jiang C. CD-MPM: continuum damage material point methods for dynamic fracture animation. ACM Trans Graph 2019;38(4):119.
[55] Lin J, Wu Y, Huang TS. Modeling the constraints of human hand motion. In: Proceeding Workshop on Human Motion; 2000 Dec 7–8; Austin, TX, USA. New York City: IEEE; 2000. p. 121–26.
[56] Lee BW, Shin H. Feasibility study of sitting posture monitoring based on piezoresistive conductive film-based flexible force sensor. IEEE Sensors J 2016;16(1):15–6.
[57] Leap motion controller [Internet]. Mountain View: ultraleap; [cited 2023 Jan 5]. Available from: https://www.ultraleap.com/product/leap-motion-controller/.
[58] Intel® RealSense™ Technology [Internet]. Santa Clara: Intel; [cited 2023 Jan 5]. Available from: https://www.intel.com/content/www/us/en/architecture-and-technology/realsense-overview.html.
[59] Feix T, Romero J, Schmiedmayer HB, Dollar AM, Kragic D. The GRASP Taxonomy of human grasp types. IEEE Trans Hum Mach Syst 2016;46(1):66–77.
[60] Liu T, Liu Z, Jiao Z, Zhu Y, Zhu SC. Synthesizing diverse and physically stable grasps with arbitrary hand structures using differentiable force closure estimator. IEEE Robot Autom Lett 2022;7(1):470–7.
[61] Zienkiewicz OC, Taylor RL. The finite element method, volume 2: solid mechanics. 5th ed. Oxford: Butterworth-Heinemann; 2000.
[62] Li M. Robust and accurate simulation of elastodynamics and contact dissertation. Pennsylvania: University of Pennsylvania; 2020.
[63] Li M, Kaufman DM, Jiang C. Codimensional incremental potential contact. ACM Trans Graph 2021;40(4):170.
[64] Fang Y, Li M, Jiang C, Kaufman DM. Guaranteed globally injective 3D deformation processing. ACM Trans Graph 2021;40(4):75.
[65] Ferguson Z, Li M, Schneider T, Gil-Ureta F, Langlois T, Jiang C, et al. Intersection-free rigid body dynamics. ACM Trans Graph 2021;40(4):183.

[66] Lan L, Yang Y, Kaufman DM, Yao J, Li M, Jiang C. Medial IPC: accelerated incremental potential contact with medial elastics. ACM Trans Graph 2021;40(4):158.
[67] Zhao Y, Choo J, Jiang Y, Li M, Jiang C, Soga K. A barrier method for frictional contact on embedded interfaces. 2021. arXiv:2107.05814.
[68] Li M, Gao M, Langlois T, Jiang C, Kaufman DM. Decomposed optimization time integrator for large-step elastodynamics. ACM Trans Graph 2019;38(4):70.
[69] Wang X, Li M, Fang Y, Zhang X, Gao M, Tang M, et al. Hierarchical optimization time integration for CFL-rate MPM stepping. ACM Trans Graph 2020;39(3):21.
[70] Nocedal J, Wright S. Numerical optimization. Berlin: Springer Science & Business Media; 2006.
[71] Hegemann J, Jiang C, Schroeder C, Teran JM. A level set method for ductile fracture. In: Proceedings of ACM SIGGRAPH/Eurographics Symposium on Computer Animation (SCA); 2013 Jul 19–21; Anaheim, CA, USA. New York City: Association for Computing Machinery (ACM); 2013. p. 193–202.
[72] Bourne M. Food texture and viscosity: concept and measurement. Amsterdam: Elsevier; 2002.
[73] Williams SH, Wright BW, Truong V, Daubert CR, Vinyard CJ. Mechanical properties of foods used in experimental studies of primate masticatory function. Am J Primatol 2005;67(3):329–46.
[74] Kiani M, Maghsoudi H, Minaei S. Determination of Poisson's ratio and Young's modulus of red bean grains. J Food Process Eng 2011;34(5):1573–83.
[75] Edmonds M, Gao F, Xie X, Liu H, Qi S, Zhu Y, et al. Feeling the force: integrating force and pose for fluent discovery through imitation learning to open medicine bottles. In: Proceedings of 2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2017); 2017 Sep 24–28; Vancouver, BC, Canada. New York City: IEEE; 2017. p. 3530–7.
[76] Xie X, Li C, Zhang C, Zhu Y, Zhu SC. Learning virtual grasp with failed demonstrations via Bayesian inverse reinforcement learning. In: Proceedings of 2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2019); 2019 Nov 3–8; Macao, China. New York City: IEEE; 2019. p. 1812–7.
[77] Rautaray SS, Agrawal A. Vision based hand gesture recognition for human computer interaction: a survey. Artif Intell Rev 2015;43(1):1–54.
[78] Dautenhahn K, Nehaniv CL. Imitation in animals and artifacts. Cambridge: MIT Press; 2002.
[79] Kubricht HKJ, Lu H. Intuitive physics: current research and controversies. Trends Cogn Sci 2017;21(10):749–59.
[80] Spelke ES. What babies know: core knowledge and composition, volume 1. Oxford: Oxford University Press; 2022.
[81] Spelke ES, Kinzler KD. Core knowledge. Dev Sci 2007;10(1):89–96.
[82] Zhu Y, Gao T, Fan L, Huang S, Edmonds M, Liu H, et al. Dark, beyond deep: a paradigm shift to cognitive AI with humanlike common sense. Engineering 2020;6(3):310–45.
[83] Zhang Z, Jiao Z, Wang W, Zhu Y, Zhu SC, Liu H. Understanding physical effects for effective tool-use. IEEE Robot Autom Lett 2022;7(4):9469–76.
[84] Li P, Liu T, Li Y, Geng Y, Zhu Y, Yang Y, et al. GenDexGrasp: generalizable dexterous grasping. 2022. arXiv:2210.00722.
[85] Zhu Y, Zhao Y, Zhu SC. Understanding tools: task-oriented object modeling, learning and recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2015); 2015 Jun 7–12; Boston, MA, USA. New York City: IEEE; 2015. p. 2855–64.
[86] Han M, Zhang Z, Jiao Z, Xie X, Zhu Y, Zhu SC, et al. Scene reconstruction with functional objects for robot autonomy. Int J Comput Vis 2022;130(12):2940–61.
[87] Han M, Zhang Z, Jiao Z, Xie X, Zhu Y, Zhu SC, et al. Reconstructing interactive 3D scene by panoptic mapping and cad model alignments. In: Proceedings of 2021 IEEE International Conference on Robotics and Automation (ICRA 2021); 2021 May 30–Jun 5; Xi'an, China. IEEE; 2021. p. 12199–206.
[88] Chen Y, Huang S, Yuan T, Zhu Y, Qi S,Zhu SC. Holistic++ scene understanding: single-view 3D holistic scene parsing and human pose estimation with human–object interaction and physical commonsense. In: Proceedings of 2019 IEEE/CVF International Conference on Computer Vision (ICCV 2019); 2019 Oct 27–Nov 2; Seoul, Republic of Korea. New York City: IEEE; 2019. p. 8647–56.
[89] Huang S, Qi S, Xiao Y, Zhu Y, Wu YN, Zhu SC. Cooperative holistic scene understanding: unifying 3D object, layout and camera pose estimation. In: Proceedings of Proceedings of the 32nd International Conference on Neural Information Processing Systems (NeurIPS 2018); 2018 Dec 3–8; Montréal, QC, Canada. Red Hook: Curran Associates Inc.; 2018. p. 206–17.
[90] Huang S, Qi S, Zhu Y, Xiao Y, Xu Y, Zhu SC. Holistic 3D scene parsing and reconstruction from a single RGB image. In: Proceedings of 2018 15th European Conference on Computer Vision (ECCV 2018); 2018 Sep 14–18; Munich, Germany. Berlin: Springer; 2018. p. 194–211.
[91] Li C, Liang W, Quigley C, Zhao Y, Yu LF. Earthquake safety training through virtual drills. IEEE Trans Vis Comput Graph 2017;23(4):1275–84.
[92] Zhu Y, Jiang C, Zhao Y, Terzopoulos D, Zhu SC. Inferring forces and learning human utilities from videos. In: Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2016); 2016 Jun 27–30; Las Vegas, NV, USA. New York City: IEEE; 2016. p. 3823–33.
[93] Zheng B, Zhao Y, Yu J, Ikeuchi K, Zhu SC. Scene understanding by reasoning stability and safety. Int J Comput Vis 2015;112(2):221–38.
[94] Zheng B, Zhao Y, Yu JC, Ikeuchi K, Zhu SC. Beyond point clouds: scene understanding by reasoning geometry and physics. In: Proceedings of the 2013 IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2013); 2013 Jun 23–28; Portland, OR, USA. New York City: IEEE; 2013. p. 3127–34.
[95] Jiao Z, Zhang Z, Wang W, Han D, Zhu SC, Zhu Y, et al. Efficient task planning for mobile manipulation: a virtual kinematic chain perspective. In: Proceedings of 2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2021); 2021 Sep 27–Oct 1; Prague, Czech Republic. New York City: IEEE; 2021. p. 8288–94.
[96] Jiao Z, Zhang Z, Jiang X, Han D, Zhu SC, Zhu Y, et al. Consolidating kinematic models to promote coordinated mobile manipulations. In: Proceedings of 2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2021); 2021 Sep 27–Oct 1; Prague, Czech Republic. New York City: IEEE; 2021. p. 979–85.
[97] Jiao Z, Niu Y, Zhang Z, Zhu SC, Zhu Y, Liu H. Sequential Manipulation Planning on Scene Graph. In: Proceedings of 2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2022); 2022 Oct 23–27; Kyoto, Japan. New York City: IEEE; 2022. p. 8203–10.
[98] Taheri O, Ghorbani N. Black MJ, Tzionas D. GRAB: a dataset of whole-body human grasping of objects. In: Proceedings of 16th European Conference on Computer Vision (ECCV 2020); 2020 Aug 23–28, Glasgow, UK. Berlin: Springer; 2020. p. 581–600.
[99] Wang Z, Chen Y, Liu T, Zhu Y, Liang W, Huang S. HUMANISE: language-conditioned human motion generation in 3D scenes. In: Proceedings of 36th Conference on Neural Information Processing Systems (NeurIPS 2022); 2022 Nov 28–Dec 9; New Orleans, LA, USA. Red Hook: Curran Associates Inc.; 2022.
[100] Jiang N, Liu T, Cao Z, Cui J, Chen Y, Wang H, et al. CHAIRS: towards full-body articulated human–object interaction. 2022. arXiv:2212.10621.
[101] Jia B, Chen Y, Huang S, Zhu Y, Zhu SC. LEMMA: a multi-view dataset for learning multi-agent multi-task activities. In: Proceedings of European Conference on Computer Vision (ECCV 2020); 2020 Aug 23–28; Glasgow, UK. Berlin: Springer; 2020. p. 1–7.