FSGlove: An Inertial-Based Hand Tracking System with Shape-Aware Calibration

Yutong Li^{1,*}, Jieyi Zhang^{1,*}, Wengiang Xu¹, Tutian Tang¹ and Cewu Lu^{1,†}

Abstract-Accurate hand motion capture (MoCap) is vital for applications in robotics, virtual reality, and biomechanics, yet existing systems face limitations in capturing high-degree-offreedom (DoF) joint kinematics and personalized hand shape. Commercial gloves offer up to 21 DoFs, which are insufficient for complex manipulations while neglecting shape variations that are critical for contact-rich tasks. We present FSGlove, an inertial-based system that simultaneously tracks up to 48 DoFs and reconstructs personalized hand shapes via DiffHCal, a novel calibration method. Each finger joint and the dorsum are equipped with IMUs, enabling high-resolution motion sensing. DiffHCal integrates with the parametric MANO model through differentiable optimization, resolving joint kinematics, shape parameters, and sensor misalignment during a single streamlined calibration. The system achieves state-of-the-art accuracy, with joint angle errors of less than 2.7°, and outperforms commercial alternatives in shape reconstruction and contact fidelity. FSGlove's open-source hardware and software design ensures compatibility with current VR and robotics ecosystems, while its ability to capture subtle motions (e.g., fingertip rubbing) bridges the gap between human dexterity and robotic imitation. Evaluated against Nokov optical MoCap, FSGlove advances hand tracking by unifying the kinematic and contact fidelity. Hardware design, software, and more results are available at: https://sites.google.com/view/fsglove.

I. INTRODUCTION

The human hand, a remarkably dexterous end-effector capable of executing intricate tasks, serves as both the inspiration and benchmark for robotic hand design and manipulation research. Accurate motion capture of the hand during manipulation is critical for collecting data essential to diverse downstream applications, including hand pose estimation, teleoperation, and imitation learning. Achieving high-fidelity hand motion capture hinges on two factors: joint kinematics and hand shape modeling.

Currently, the most comprehensive commercially available hand MoCap glove [1] can capture at most 21 degrees of freedom (DoFs). However, complex manipulations, such as thumb-index fingertip rubbing, are still beyond the ability to capture, as they require additional torsional DoFs in the proximal or middle phalanges that are often overlooked (Fig. 1), but recently, some dexterous robotic hands [2] tend to support control such DoFs. Similarly, hand shape



Fig. 1: Hand shape variation. (a) The variation in hand shapes across different human subjects and a robotic hand. (b) Thumb-index fingertip rubbing requires a slight twist DoF. (c), (d) Motion transfer often fails when adapted between hands with heterogeneous shapes, even when all the joints are mapped.

variations, driven by differences in bone length and softtissue composition, demand personalized models for accurate motion reconstruction, particularly during contact-rich tasks involving objects or self-interaction. Existing hand motion capture systems, whether commercial [3], [4] or researchoriented [5], [6], [7], focus primarily on joint angle estimation with limited DoFs and neglect the shape variations. This oversight hinders the transfer of in-manipulation motions to virtual avatars or anthropomorphic robotic hands, where both kinematic and contact fidelity are paramount (Fig. 1).

To bridge this gap, we present FSGlove, a novel inertialbased hand tracking system simultaneously capturing up to 48-DoF joint kinematics and personalized hand shape. Each finger joint and the dorsal region are instrumented with inertial measurement units (IMUs), enabling high-resolution motion sensing. A key innovation is DiffHCal, a shapeaware calibration method that embeds personalized handshape estimation into a streamlined calibration protocol. By leveraging the parametric MANO hand model [8], DiffHCal aligns captured sensor data to MANO's joint and shape parameters through a differentiable optimization process. This framework uses a series of predefined reference poses to simultaneously resolve skeletal joint parameters, shape parameters, and manual installation errors in a single optimization process. Moreover, this process integrates seamlessly into standard glove calibration workflows, requiring no additional steps compared with commercial systems while delivering

^{*} indicates equal contribution.

^{1{}yutong, yi_eagle, vinjohn, tttang, lucewu}@sjtu.edu.cn. Yutong Li, Jieyi Zhang, Wenqiang Xu, and Tutian Tang are with School of Electronic Information and Electrical Engineering, Shanghai Jiao Tong University, Shanghai, China, and also with the Meta Robotics Institute, Shanghai Jiao Tong University, Shanghai, China. Cewu Lu is the corresponding author, a member of Qing Yuan Research Institute and MoE Key Lab of Artificial Intelligence, AI Institute, Shanghai Jiao Tong University, China.

contact-consistent hand models. To democratize access, we open-source the hardware design, low-level drivers for sensor calibration, and high-level interfaces for integration with motion capture ecosystems (e.g., OptiTrack, HTC Vive), enabling plug-and-play compatibility with robotics and VR frameworks.

We evaluate FSGlove across four metrics: (1) joint angle accuracy ($\leq 2.7^{\circ}$), (2) shape reconstruction precision (mean mesh error ≤ 3.6 mm), (3) fingertip tracking error (mean positional error ≤ 16 mm), and (4) hand-object interaction fidelity (mean positional error ≤ 20 mm), benchmarking against commercial glove systems. Despite its low-cost design, FSGlove achieves state-of-the-art performance in shape-aware tracking, surpassing commercial alternatives (e.g., Manus Metaglove Pro, VRTRIX) in contact-rich manipulation tasks.

We summarize our contributions as follows:

- FSGlove, the first open-source, high-DoF (up to 48 DoFs) data glove integrating inertial sensing with shape-aware calibration.
- DiffHCal, a differentiable calibration framework that infers personalized hand shape during standard wear-time procedures.
- Comprehensive validation demonstrating high-precision accuracy in joint measurement and shape reconstruction.

II. RELATED WORKS

A. Data Glove Systems

The development of data gloves dates back to the 1970s. Since then, various types based on different sensors have emerged, each with distinct pros and cons. The early Sayre Glove in 1977 [9] used flexible tubes and photocells. It could measure finger bending but had few sensors, was hardwired and cumbersome, and had limited applications. In the 1980s, the Data Glove [10] was a significant advancement. It used plastic tubes and light-based sensors for joint angle measurement, followed by the Power Glove [11] and the Super Glove [12]. These gloves could measure finger joint bending but the cloth support restricted hand movement, affecting measurement accuracy. They also needed complex calibration, and some had problems such as non-linear mapping and sampling limitations [12].

MemGlove[13] combined resistive and fluidic sensors for hand pose reconstruction. However, it featured a complex fabrication and sensor integration process. Gloves with flex sensors [14] were inexpensive and easy to use for measuring finger flexion. However, the sensors were sensitive to temperature and vulnerable to wear and tear. In addition, the large size of the sensors limited the degrees of freedom they could capture.

With advancements in microelectronics, gloves equipped with MEMS-based inertial measurement unit (IMU) sensors have emerged. The standardized production of sensor chips has significantly reduced costs, enabling mass production of these gloves. These gloves [15], [16] leverage IMU characteristics to provide superior dynamic responses, enabling comprehensive capture of hand and arm movements. [6] introduced a data glove prototype incorporating flexible printed circuits (FPC) and proposed the use of angle sensors to verify IMU accuracy. Similarly, [17] also presented a modular IMU-based data glove design. However, these approaches neglected adaptability to varying hand sizes and lacked optimization for machine learning applications. Nonetheless, these sensors face calibration challenges [16] and occasional susceptibility to external magnetic field interference.

B. Glove Calibration Methods

To alleviate or address problems caused by glove sensors, data glove calibration research has seen continuous development in recent years. Chou et al. [18] used linear regression for Cyberglove calibration, relying on a vision system to handle the complex relationship between sensor readings and joint angles. Kahlesz et al. [19] modeled sensor cross-couplings for high-degree DoF data gloves like the Cyberglove, achieving visually appealing calibrations without extra hardware. Sun et al. [20] applied genetic algorithms to calibrate exoskeleton data gloves, accounting for hand-size and wearing-position variations. Zhou et al. [21] simplified CyberGlove calibration in virtual rehabilitation via artificial neural networks, benefiting disabled patients. Connolly et al. [22] improved the accuracy and usability of the 5DT Data Glove 14 Ultra using neural networks, eliminating the need for traditional calibration for those with limited joint mobility.

III. SYSTEM ARCHITECTURE AND DESIGN

For both robotics and motion capture applications, stable operation and real-time processing of data from multiple sensors are essential. To meet this requirement, we have developed the distributed motion capture system shown in Fig. 2. The system comprises three hardware-independent subsystems: (1) the FSGlove for finger kinematics recording, (2) a dorsal tracker for global hand translation, and (3) a data acquisition suite for sensor fusion and visualization. These components communicate via Ethernet or Wi-Fi with synchronized clocks. We detail the FSGlove hardware in Sec. III-A, the dorsal tracker in Sec. III-B, and the data acquisition pipeline in Sec. III-C.

A. Hardware Setup of FSGlove

a) Topology and Mechanical Design: Human finger joints function as hinge-like constraints, permitting rotations between hierarchically connected phalanxes (proximal \rightarrow middle \rightarrow distal). To capture this kinematic chain, we deploy three IMUs per finger, positioned dorsally on the proximal, middle, and distal phalanxes, supplemented by a dorsummounted IMU for global orientation. This forms a 16-IMU network (5 fingers × 3 + 1 dorsum).

To minimize wiring complexity while preserving dexterity, IMUs are soldered onto a custom FPC. The FPC is bonded to a stretchable nylon glove using 3D-printed mounts, which secure sensors against skin slippage during dynamic motions. A dorsum-mounted enclosure houses the single-board



Fig. 2: **System Architecture.** The system gathers IMU sensor data via an SBC and then sends them from the palm tracker to a host computer. The computer calibrates and models the hand shape, aligns the data, and outputs a final hand-and-object mesh.



Fig. 3: **Glove hardware design.** The Raspberry Pi Zero 2W serves as the system's core, receiving sensor data from a custom UART-USB daughter board over a USB bridge. The daughter board collects IMU data via a flex-printed circuit, and the IMUs are attached to the glove surface.



Fig. 4: **FSGlove's board compatibility.** (a) Basic FSGlove without a dorsal tracker, capturing only finger kinematics. (b) FSGlove is equipped with an optical dorsal tracker and tapped FPC, which is integrated with the Nokov MoCap system. (c) FSGlove is equipped with an HTC Vive Tracker, integrated with the Vive VR system.

computer (SBC) and USB-UART daughter board, ensuring synchronous movement with the hand's rigid metacarpal structure.

b) *IMU Selection:* The IMU dimensions must accommodate finger phalanx size while providing high-accuracy orientation estimation. We prioritize compact form factor, estimation precision (static/dynamic), refresh rate, and cost efficiency. Our implementation employs the HI229 module [23] based on the BNO055 chip, which delivers world-frame-referenced fused orientation estimates via onboard sensor fusion. As specified in the manufacturer datasheet, this module achieves 0.8° static angular error and 2.5° dynamic error at 100-400 Hz sample rates – performance comparable with commercial data glove IMUs. For example, VRTRIX datagloves[4] can achieve $0.5-2.0^{\circ}$ angular error at 180Hz sample rate.

c) Single-Board Computer (SBC): The SBC should be able to process 16 IMU data streams with wireless support. We select the Raspberry Pi 2W (Pi2W) for its quad-core 1.0GHz ARM CPU and integrated Wi-Fi, which satisfies real-time data aggregation and transmission requirements. As a widely available off-the-shelf component, the Pi2W ensures system replicability and simplifies maintenance.

d) USB-UART Daughter Board: IMU data should be sent to SBC through UART ports. To drive 16 IMUs via UART ports, we designed a daughter board integrating a USB hub and dual CH9344 8-Port USB-UART bridge chips to expand the connectivity to 16 UART ports while maintaining 480 Mbps uplink bandwidth. A custom USB bridge connects the daughter board to the Pi2W's USB controller, and all components are housed in a 3D-printed enclosure to prevent electrical shorts. This design resolves the port scarcity inherent to SBCs without compromising signal integrity.

e) Battery System for Mobility: A 1100 mAh Li-ion battery with charge-management circuitry powers the Pi2W via its 40-pin GPIO header, enabling untethered operation. Despite the high power demands of 16 IMUs and Wi-Fi communication, the glove can operate continuously for 2 hours. This balance between runtime and glove weight (110g) ensures usability during prolonged manipulation tasks.

f) Cost Analysis: The total bill-of-materials cost for the 16-IMU glove configuration is \$426, including assembly expenses (detailed breakdown in Supplementary). IMUs account for 75% of this cost (\$320), suggesting cost reductions are achievable in applications requiring fewer sensors—consistent with commercial glove designs. For such scenarios, functionally equivalent systems can be realized at substantially lower costs while retaining compatibility with our modular architecture.

B. Dorsal Tracker

While IMUs capture hand joint orientations, global translational motion requires complementary tracking. We adopt the Nokov optical system to resolve 6DoF dorsal poses. As is shown in Fig. 4, we design a 3D-printed tree-like structure hosting multiple infrared reflective (IR) markers, to ensure the glove is tracked stably with the optical system. Additionally, we apply tape to shield the FPC, thereby reducing infrared reflection noise.

C. Data Acquisition Suite

The data acquisition suite integrates synchronized streams from the FSGlove and the dorsal tracker. We detail four critical design aspects below.

a) Communication Protocols: We adopt gRPC for data transmission, leveraging its low-latency bidirectional communication and stream abstraction capabilities. Integration of the Protobuf serialization protocol eliminates custom communication protocol development, enhancing maintainability compared to naive TCP implementations. For dorsal tracking, a middleware layer facilitates interoperability with heterogeneous tracking systems, utilizing inter-process communication (IPC) to minimize latency and computational overhead.

b) Time Synchronization: The distributed design requires a synchronized clock of different sub-systems. In this pipeline, the SBC synchronizes its system clock with a PCbased time-server using the popular Network Time Protocol. Given the limited crystal quality and inherent inaccuracies of SBC system clocks, this synchronization is performed periodically to maintain alignment with the PC's time. The Nokov system employs custom protocols for synchronization, but for other systems, the Precision Time Protocol [24] is recommended, especially for devices with modern network controllers. The 16 onboard IMUs are synchronized by the time sync module based on the timestamped arrival of data packets. To account for potential variations in message arrival times, the synchronization process utilizes an approximate policy derived from the message filter of the Robot Operating System (ROS).

c) The Data Flow: The SBC and PC communicate over Wi-Fi. During initialization, the SBC loads USB drivers and processes raw IMU byte streams through a parser module, converting them into timestamped structured messages. To optimize the trade-off between battery consumption, computational load, and motion fidelity, IMUs are sampled at 100 Hz. A time synchronization module aligns data from all 16 IMUs into unified measurement entries containing triaxial acceleration, angular velocity, and fused orientation data. These entries are streamed to the PC via gRPC, where a calibration module integrates dorsal tracker poses with the shape-aware calibration algorithm to resolve the final hand pose. Algorithmic details are elaborated in Sec. IV-A.

d) Delay Analysis: System latency measurements quantify a total glove delay of 24 ms, predominantly caused by Wi-Fi channel congestion at the SBC. The Nokov tracker exhibits significantly lower latency (1 ms) thanks to direct wired connectivity. Middleware and memory operations introduce negligible overhead (< 0.5 ms) on modern PCs with multicore CPUs and hierarchical cache architectures. Reconstruction algorithm efficiency further influences endto-end latency, with GPU-accelerated processing (NVIDIA 2080 Ti) enabling real-time visualization at 25 Hz, yielding a display latency of 40 ms.

IV. MOCAP DATA ALIGNMENT AND SHAPE-AWARE CALIBRATION

In the FSGlove-based MoCap system, we need to convert the raw sensor readings to hand motion data. We propose a unified approach, DiffHCal, to align and calibrate the data from the MoCap system and consider both joints and shapes. DiffHCal is built based on the MANO hand model [8], which is widely utilized in many hand-related tasks and has a rich ecosystem. Importantly, it parameterized both the joint and shape of the hand.

A. Pose Alignment and Calibration

The raw IMU data captured by the glove does not represent joint angles but the phalange rotations. Therefore, we need to first convert them to joint angles. Then, considering the glove's wearing, the IMU may have different relative poses to the corresponding finger phalange for each glove's wearing process. Such installation poses should be calibrated each time the glove is worn.

a) Pose Alignment: We first establish a virtual world coordinate system \mathcal{M} for MANO. IMU has a uniform West-North-Up (WNU) coordinate system as the world coordinate system \mathcal{W} . IMUs are indexed according to the finger link ID on MANO. Ideally, we assume that each IMU is installed and aligned with the link frame perfectly. $R_i^{\mathcal{W}}$ is *i*-th IMU's reading while $R_i^{\mathcal{M}}$ represents the rotation of *i*-th finger link in the M coordinate system. Then, rotation from the WNU world coordinate system to the MANO's world coordinate system is A:

$$R_i^{\mathcal{M}} = A R_i^{\mathcal{W}}.$$
 (1)

By defining the joint index with its parent link's index, in MANO's convention, for each finger, joint i connects link i and link i - 1. Therefore, the joint angle can be calculated by:

$$\theta_i = R_i^{\mathcal{M}} (R_{i-1}^{\mathcal{M}})^{-1}.$$
(2)

b) Pose Calibration: In Eq. 1 and Eq. 2, $R_i^{\mathcal{W}}$ is read from IMU sensor, $R_i^{\mathcal{M}}$ is read from the virtual environment where the MANO exists (*i.e.*, a 3D simulator, such as RFUniverse [25]). All we need to know is the transformation matrix A.

The challenge in calculating A is how a human in the real world can pose the hand the same way as the virtual hand. Since A is the same for all the finger links, one finger link pose pair is all we need.

However, in practice, the IMU cannot be perfectly aligned with the link frame, and its orientation changes each time the glove is worn because it is flexible. Thus, considering such installation error, Eq. 1 should be:

$$R_i^{\mathcal{M}}C_i = A R_i^{\mathcal{W}},\tag{3}$$

 C_i is the correction rotation indicating the installation error from *i*-th link frame.



Fig. 5: **Calibration Poses.** (a) The pose calibration process involves only three simple poses. (b) The shape calibration process requires each fingertip to touch the thumb tip.

To solve both A and C_i , at least two sets of standard reference hand pose should be defined, and the least square method can be applied.

In practice, we define three standard reference poses: the rest pose, the x-axis rotation pose, and the y-axis rotation pose. As shown in Fig. 5.(a), The rest pose is when flatting the hand palm and face downward onto a surface, and spreading all fingers as wide as possible. The x-axis rotation pose is when we rotate the wrist of the rest pose around the x-axis (keeping the fingers still and orienting the palm vertically). The y-axis rotation pose is when we rotate the wrist of the x-axis (keeping the fingers still and orienting the palm vertically). The y-axis rotation pose around the y-axis (keeping the fingers still while moving the palm in a vertical plane).

B. Shape Calibration

When two people try to manipulate an object with the same hand pose, the one with the bigger hand is already in contact with the object, while the one with the smaller hand is not. Failing to consider the shape variation will cause severe problems when downstream tasks (e.g., VR/AR, manipulation data collection) involve object interaction. Therefore, we also need to calibrate the shape size.

To personalize different human hand shapes, we use the MANO shape parameter vector β . Given pose parameters θ and shape parameters β , the hand mesh model is defined as

$$(V, E) = M(\theta, \beta), \tag{4}$$

where V denotes the vertex set, E denotes the edge set.

One way to personalize the hand shape is to scan the hand with a glove and apply the hand rig to the scanned model. It would require considerable modeling time and is not very practical from the user end. Another way is to utilize a personalized hand reconstruction method such as [26], but such a method is not trained on the hand with gloves.

Considering that shape variation mostly influences the contact states, we propose calibrating the hand shape using a few touch-based standard reference poses. In the standard pose, we will define a contact state $c_{jk} = (v_j, v_k)$, the contact set $C = \{c_{jk}\}, v_j$ is the *j*-th vertex. As shown in Fig. 5. (b), in a thumb-index pinch pose, if the thumb is in contact with the index finger, 744-th vertex and 320-th

TABLE I: **Comparison of hand glove motion capture solutions.** Our glove offers the lowest cost while maintaining the sampling rate and pose DoFs.

Solution	Technology	Sampling Rate (Hz)	Price
Ours (FSGlove)	IMU (16 unit)	100	\$426
VRTRIX Pro	IMU (6 unit)	180	\$1,400
Meta Quest 3	Camera	60	\$499
Manus Metaglove	Magnetic Sensor	120	\$5,250

vertex should be in contact, where indices are of MANO's convention. Given the predefined contact relationship in the reference poses, and humans try to replicate the reference pose in the real world, we apply the measured hand pose to the MANO hand and define the shape error metric as:

$$E_{shape}(\beta) = \sum_{c_{jk} \in \mathcal{C}} \|v_j(\theta, \beta) - v_k(\theta, \beta)\|_2^2.$$
(5)

Since M in Eq. 4 is differentiable [8], minimizing E_{shape} to find the optimal β is also differentiable.

Considering most manipulation applications are graspingbased, optimizing the contact states on the fingertips would suffice for such scenarios. It is the default setting for our experiments. If more complex manipulations are considered, such as in-hand manipulation, the contact states on another part of the hand could also be considered, which is trivial to extend with the same optimization scheme.

C. Alignment with the Dorsal Tracking System

Finally, the wrist pose of the personalized MANO must be obtained. We mentioned before that in the FSGlove-based MoCap system, the dorsal rotation is obtained from IMU, and the dorsal translation is obtained from the external dorsal tracker. The alignment and calibration process is similar to the pose part. The main difference is that we should calculate the transformation matrix A' between the dorsal tracker's world coordinate system and MANO's.

To note, we can use the same reference hand poses as the ones described in Sec. IV-A to accomplish the alignment process. Thus, we can simultaneously accomplish the pose alignment process for the hand and wrist if necessary. In this way, all the calculations for hand pose, shape, and wrist pose are differentiable.

V. EVALUATION EXPERIMENTS

To evaluate the performance of these gloves, we designed a series of experiments based on common research and application scenarios and selected multiple different hand pose estimation technologies for comparative study. They include the VRTRIX MoCap glove, which also uses an IMU solution, the Meta Quest3, which adopts a vision-based hand recognition algorithm, and the Manus Metaglove Pro based on magnetic sensing, which claims to be the most precise hand tracking solution. Details of their information can be found in Tab. I.



Fig. 6: Settings of single joint error experiments. We use optical system as ground truth to test the single joint error.



Fig. 7: **Results of single joint measurement.** Our system's angle output closely matches the measurements from the Nokov MoCap system.

A. Single Joint Measurement

According to our specific use case, we need to validate the accuracy of joint pose alignment using a pair of IMUs. We have designed a 3D-printed base plate to model scenarios of finger movement. As illustrated in Fig.6, the device is divided into two sections, A and B, connected via a rotating axis. Each section has designated positions for mounting motion capture marker balls, facilitating the capture of sections A and B's postures using the Nokov system and enabling the calculation of the true rotation angle of the axis. With Nokov's tracking accuracy at 0.5 mm, at a scale of 100 mm, the measurement accuracy for rotation reaches 0.3 degrees, which meets our requirements. In our experiments, we use the output from the Nokov system as the benchmark to gauge the error in pose estimation provided by the IMU system. As the results shown in Fig.7, the bias of the IMU's pose estimation is $\pm 2.7^{\circ}$, with a standard deviation of $\pm 1.8^{\circ}$. The non-linearity of the sensor is 0.7%. Furthermore, its accuracy does not fluctuate with changes in joint angles. We can deem such an IMU system is reliable.

Method \downarrow	Distance (mm)						
	Index	Middle	Ring	Little	Average		
Ours	7.2	16.5	12.0	27.1	15.7		
Meta Quest3	11.0	21.5	18.9	22.3	19.6		
VRTRIX	20.9	14.4	25.3	15.1	18.9		
MANUS	28.0	32.3	36.7	35.7	33.2		

TABLE II: **Results of fingertip pinch tracking.** Our method achieves the best average error. Notably, it attains the lowest error for the index and ring fingers, confirming its superior accuracy in most finger measurements.



Fig. 8: **Quantitative results of shape reconstruction.** Our solution demonstrates comparable performance to the Quest3 and outperforms other commercially available gloves.

B. Shape Reconstruction

As a data glove designed for hand motion reconstruction, we are particularly interested in its fundamental shape reconstruction capabilities. To conduct this experiment, a Photoneo MotionCam depth camera was mounted on a tripod and oriented toward a flat white wall, with the space between the camera and the wall cleared to ensure unobstructed imaging. A volunteer wearing the glove performed a series of complex and diverse hand poses. During this process, the Photoneo camera captured a high-precision grayscale depth image of the volunteer's hand while the glove recorded its reconstruction results concurrently.

To compare the reconstruction results with the highprecision depth images, the GroundedSAM model [27] was employed to apply a "hand" query to the depth image, enabling precise extraction of the volunteer's hand. The extracted regions were then converted into a point cloud and compared with the glove's reconstruction results. To account for the Photoneo camera's output of a partial point cloud, we use the unidirectional Chamfer Distance from the partial point cloud P to the reconstructed hand mesh vertices V as the error metric:

$$E_{sr} = \frac{1}{|P|} \sum_{p \in P} \min_{v \in V} \|p - v\|_2^2$$

The experiment was repeated 10 times for each pose, and the results were averaged. As summarized in Fig. 8, our solution demonstrates comparable performance to the Quest3 and outperforms other commercially available gloves.

However, we find that the Meta Quest3 performs comparably or even better in fingertip pinching. We attribute this to two main factors. First, Quest3 uses visual information to reconstruct a hand mesh, thereby capturing more precise geometric details when subtle contact actions are visible.



Fig. 9: Qualitative results of shape reconstruction.

		Test Result (mm)							Total Avg.		
Method ↓	#1		#2		#3		#4		#5		-
	Avg.	Std.	Avg.	Std.	Avg.	Std.	Avg.	Std.	Avg.	Std.	
Ours	22.6	5.5	16.5	4.2	17.9	2.9	26.6	2.0	17.5	3.3	20.2
Meta Quest3	69.7	13.0	17.7	1.3	18.0	4.8	16.7	11.1	18.1	9.2	28.0
VRTRIX	41.9	6.4	14.4	2.3	19.1	3.6	16.2	2.3	30.8	8.2	24.5
MANUS	28.0	7.3	19.8	3.9	26.7	4.6	25.7	4.0	33.2	12.3	26.7

TABLE III: **Quantitative results of Hand-Object Interaction Consistency Experiment** As can be observed, our proposed method outperforms other approaches on most objects. The relatively lower scores of MANUS and VRTRIX are largely attributable to unnatural hand motions induced by external forces during the interaction, which are consequently filtered by their built-in algorithms.

Second, the finger-mounted dorsal IMUs can collide with each other, especially during challenging poses like thumbpinky contact, causing positional drift and reducing accuracy.

C. Fingertip Pinch Tracking

Motion capture gloves are commonly utilized in virtual reality (VR). For example, [28] adopted a commercial motion capture glove in VR to generate a dataset that trains a grasp policy for robot arms. Therefore, minimizing contact error between fingertips is critical, as it directly impacts the accuracy of essential operations such as grasping. To address this, we designed an experiment to quantify fingertip contact error in the Pinch pose. After properly calibrating the glove, participants performed a series of instructed gestures, sequentially bringing the thumb into contact with the other four fingertips. During the experiment, wrist rotation was randomly applied to simulate real-world operating conditions. For each Pinch pose, a 10-second sequence of distances between fingertip endpoints was recorded, and the averaged measurements were summarized in Tab. II. Our solution demonstrated marginally superior performance compared to the VRTRIX glove, whereas the MANUS glove exhibited some deviation. The Quest3's performance was notably suboptimal due to its helmet-mounted cameras' inability to capture hand poses when the user's palm faced downward.

For the *little pinching* pose, VRTRIX outperforms both our glove and other methods by a large margin. This is because

VRTRIX systematically tends to bend the thumb across the palm, regardless of the actual thumb position, as shown in Fig.9, thereby improving its performance in the evaluated metric. However, this behavior can cause the reconstructed mesh to deviate significantly from the ground truth.

D. Hand-Object Interaction Consistency

In many situations, the human hand is used to manipulate various objects. Therefore, we are very curious whether the recorded hand mesh sequences can faithfully reflect the actual physical processes that occur when gloves interact with objects. To explore this, we selected five objects: cube_large, cup, utah_teapot, stanford_rabbit, and elephant from the ContactPose [29] dataset, which provides a collection of digital models of common objects, each equipped with mounting points for marker balls.

Following the methods described in [29], we installed markers on these models to track their position and orientation changes using Nokov. We then adjusted the visualization and recording programs based on the Nokov readings. Our glove setup already includes a tracking solution for the palm, so no modifications were necessary. For VRTRIX, MANUS gloves, and Quest3, we created multiple replicas of the dorsal tracker for palm tracking. For each combination of object and glove, we had volunteers perform operations for approximately 60 seconds, recording the palm and object's position and orientation via Nokov, and hand reconstruction results via custom methods.

Our focus was on the regions of interest in the hand mesh (the tips of each finger) and the interpenetration of the hand mesh with the object mesh. We selected 30 random moments from the sequence, ensuring that all fingers were in contact with the object. For these moments, we calculated the distance from the fingertip endpoint to the nearest point on the object mesh. A lower value indicates a better reconstruction. Since all palm positions and orientations were provided by Nokov and the tracking devices, this allowed for a fair comparison. The experimental results in Tab. III indicate that our method performs slightly better than Quest3 and VRTRIX while significantly outperforming MANUS gloves.

VI. CONCLUSION AND FUTURE WORK

This work introduces FSGlove, an open-source, high-DoF (48-DoF) data glove combining inertial sensing with personalized shape calibration, addressing critical gaps in hand motion capture. By integrating densely instrumented IMUs and DiffHCal, a differentiable calibration framework, FSGlove achieves anatomically consistent hand modeling, resolving joint kinematics, shape parameters, and sensor misalignment in a unified workflow. Experiments prove its state-of-theart joint angle accuracy, robust shape reconstruction, and superior contact fidelity compared to commercial systems. The system's open design and compatibility with VR and robotics systems democratize access to high-fidelity hand tracking, enabling more possible downstream applications.

Future work includes enhancing IMU precision for subdegree tracking, extending DiffHCal to adapt to soft-tissue deformation during contact dynamically, and generalizing shape estimation to diverse hand anthropometries. Integrating tactile feedback could further improve interaction realism, while federated learning frameworks might enable personalized calibration without manual intervention. Exploring applications in surgical robotics or immersive teleoperation will validate FSGlove's scalability. By open-sourcing the design, we invite community-driven advancements, bridging the fidelity gap between human and robotic manipulation.

REFERENCES

- CyberGlove, "Cyberglove iii introduction," https://www. cyberglovesystems.com/cyberglove-iii/, 2025, accessed: March 1, 2025.
- [2] K. Shaw, A. Agarwal, and D. Pathak, "Leap hand: Low-cost, efficient, and anthropomorphic hand for robot learning," *Robotics: Science and Systems (RSS)*, 2023.
- [3] Manus, "Manus glove product introduction," https://www. manusmachina.com, 2025, accessed: March 1, 2025.
- [4] VRTrix, "Vrtrix glove product introduction," https://www.vrtrix.com, 2025, accessed: March 1, 2025.
- [5] S. Zhu, A. Stuttaford-Fowler, A. Fahmy, C. Li, and J. Sienz, "Development of a low-cost data glove using flex sensors for the robot hand teleoperation," in 2021 3rd International Symposium on Robotics & Intelligent Manufacturing Technology (ISRIMT), 2021, pp. 47–51.
- [6] B.-S. Lin, I.-J. Lee, P.-Y. Chiang, S.-Y. Huang, and C.-W. Peng, "A modular data glove system for finger and hand motion capture based on inertial sensors," *Journal of Medical and Biological Engineering*, vol. 39, pp. 532–540, 2019.
- [7] B. O'Flynn, J. Torres, J. Connolly, J. Condell, K. Curran, and P. Gardiner, "Novel smart sensor glove for arthritis rehabiliation," in 2013 IEEE International Conference on Body Sensor Networks, 2013, pp. 1–6.

- [8] J. Romero, D. Tzionas, and M. J. Black, "Embodied hands: Modeling and capturing hands and bodies together," ACM Transactions on Graphics, (Proc. SIGGRAPH Asia), vol. 36, no. 6, Nov. 2017.
- [9] D. Sturman and D. Zeltzer, "A survey of glove-based input," *IEEE Computer Graphics and Applications*, vol. 14, no. 1, pp. 30–39, 1994.
- [10] H. Eglowstein, "Reach out and touch your data," *Byte*, vol. 15, no. 7, pp. 283–290, July 1990.
- [11] D. L. Gardner, "The power glove," Des. News, vol. 45, pp. 63–68, Dec. 1989.
- [12] L. Dipietro, A. M. Sabatini, and P. Dario, "A survey of glove-based systems and their applications," *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, vol. 38, no. 4, pp. 461–482, 2008.
- [13] J. Hughes, A. Spielberg, M. Chounlakone, G. Chang, W. Matusik, and D. Rus, "A simple, inexpensive, wearable glove with hybrid resistivepressure sensors for computational sensing, proprioception, and task identification," *Advanced Intelligent Systems*, vol. 2, no. 6, p. 2000002.
- [14] S. Zhu, A. Stuttaford-Fowler, A. Fahmy, C. Li, and J. Sienz, "Development of a low-cost data glove using flex sensors for the robot hand teleoperation," in 2021 3rd International Symposium on Robotics & Intelligent Manufacturing Technology (ISRIMT), 2021, pp. 47–51.
- [15] B. Fang, F. Sun, H. Liu, and D. Guo, "A novel data glove for fingers motion capture using inertial and magnetic measurement units," in 2016 IEEE International Conference on Robotics and Biomimetics (ROBIO), 2016, pp. 2099–2104.
- [16] Q. Liu, G. Qian, W. Meng, Q. Ai, C. Yin, and Z. Fang, "A new immu-based data glove for hand motion capture with optimized sensor layout," *International Journal of Intelligent Robotics and Applications*, vol. 3, no. 1, pp. 19–32, March 2019, (c) 2019, Springer Nature Singapore Pte Ltd. This is an author produced version of a paper published in the International Journal of Intelligent Robotics and Applications. Uploaded in accordance with the publisher's self-archiving policy.
- [17] B.-S. Lin, I.-J. Lee, S.-Y. Yang, Y.-C. Lo, J. Lee, and J.-L. Chen, "Design of an Inertial-Sensor-Based Data Glove for Hand Function Evaluation," *Sensors*, vol. 18, no. 5, p. 1545, May 2018.
- [18] T.-S. Chou and A. Gadd, "Hand-eye: A vision-based approach to data glove calibration," 2000.
- [19] F. Kahlesz, G. Zachmann, and R. Klein, "visual-fidelity' dataglove calibration," in *Proceedings Computer Graphics International*, 2004., 2004, pp. 403–410.
- [20] Z. Sun, G. Bao, J. Li, and Z. Wang, "Research of dataglove calibration method based on genetic algorithms," in 2006 6th World Congress on Intelligent Control and Automation, vol. 2, 2006, pp. 9429–9433.
- [21] J. Zhou, F. Malric, and S. Shirmohammadi, "A new hand-measurement method to simplify calibration in cyberglove-based virtual rehabilitation," *IEEE Transactions on Instrumentation and Measurement*, vol. 59, no. 10, pp. 2496–2504, 2010.
- [22] J. Connolly, J. Condell, K. Curran, and P. Gardiner, "Improving data glove accuracy and usability using a neural network when measuring finger joint range of motion," *Sensors*, vol. 22, no. 6, 2022.
- [23] S. L. Tech, "Hi229 9-axis imu/vru/ahrs," https://sealandtech.com.tw/ en/product_anrot_hi229.html.
- [24] "Ieee standard for a precision clock synchronization protocol for networked measurement and control systems," *IEEE Std 1588-2019* (*Revision ofIEEE Std 1588-2008*), pp. 1–499, 2020.
- [25] H. Fu*, W. Xu*, R. Ye*, H. Xue, Z. Yu, T. Tang, Y. Li, W. Du, J. Zhang, and C. Lu, "Demonstrating rfuniverse: A multiphysics simulation platform for embodied ai," in *RSS Robotics: Science and Systems*, 2023.
- [26] K. Karunratanakul, S. Prokudin, O. Hilliges, and S. Tang, "Harp: Personalized hand reconstruction from a monocular rgb video," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2023, pp. 12802–12813.
- [27] T. Ren, S. Liu, A. Zeng, J. Lin, K. C. Li, H. Cao, J. Chen, X. Huang, Y. Chen, F. Yan, Z. Zeng, H. Zhang, F. Li, J. Yang, H. Li, Q. Jiang, and L. Zhang, "Grounded sam: Assembling open-world models for diverse visual tasks," arXiv preprint arXiv:2401.14159, 2024.
- [28] H. Xue, Y. Li, W. Xu, H. Li, D. Zheng, and C. Lu, "Unifolding: Towards sample-efficient, scalable, and generalizable robotic garment folding," in *Conference on Robot Learning*. PMLR, 2023, pp. 3321– 3341.
- [29] S. Brahmbhatt, C. Tang, C. D. Twigg, C. C. Kemp, and J. Hays, "Contactpose: A dataset of grasps with object contact and hand pose," in *16th European Conference on Computer Vision (ECCV 2020)*, 2020, pp. 361–378.