Supplementary Materials for

A Cross-Domain Orthogonal Spherical Interface for Bidirectional Haptic Human-Robot Interaction

Shoujie Li*†, Xingting Li†, Chenxin Liang†, Xuanbing Xie,
Wei Hou, Zijian Lin, Chongxi Jiao, Mingshan He, Junwei Li,
Jie Pan, Chuqiao Lyu, Xueqian Wang, Yifan Wang*†, Wenbo Ding*†
*Corresponding author. Email: shoujie.li@ntu.edu.sg; yifan.wang@ntu.edu.sg;
ding.wenbo@sz.tsinghua.edu.cn
†These authors contributed equally to this work.

This PDF file includes:

Supplementary Text 1 to 8

Figures S1 to S27

Tables S1 to S10

Captions for Movies S1 to S8

Caption for Supplementary Data S1

Other Supplementary Materials for this manuscript:

Movies S1 to S8

Supplementary Data S1

Supplementary Text

Supplementary Text 1 Principles of Spherical Design

The spherical form factor of FusionTouch was chosen to match the natural geometry of hand motion and to satisfy practical requirements for multimodal haptic interaction. We recruited 20 volunteers and collected hand-motion data during 20 representative daily manipulation tasks, yielding 240,677 pose frames. Relaxed-hand analysis showed that the palm and fingers naturally formed a curled configuration that could be approximated by a near-spherical envelope. Frame-wise minimum-enclosing-sphere analysis further showed that 98.95% of the recorded hand-motion frames had an envelope diameter smaller than 15 cm. These results indicate that a compact spherical workspace can cover most natural hand configurations involved in daily manipulation while allowing glove-free interaction.

The sphere also provides several design advantages for the FusionTouch system. First, its continuous curvature avoids the sharp edges and curvature discontinuities present in polyhedral geometries, helping reduce local stress concentration on the soft shell. Although the actual mechanical response depends on material properties, seams, inflation pressure, and boundary constraints, a smoothly curved surface provides a favorable geometry for spatially consistent contact and deformation sensing. Second, the spherical surface supports a modular petal-like fabrication strategy. The optoelectronic and electrotactile skin can be fabricated as repeated planar fan-shaped sectors and then assembled into a curved surface, reducing pattern complexity and enabling local replacement of damaged modules.

Third, the convex spherical surface is compatible with the concave posture of the relaxed hand. This geometry supports the palm and fingers without enclosing them, increases the potential contact area, and helps distribute interface pressure during prolonged use. Fourth, the spherical cavity provides a compact optical layout for internal vision. When cameras and illumination modules are placed inside the transparent shell, visible-light hand observation and near-infrared marker imaging can cover a broad interaction area. In an ideal centered configuration, the surface points have similar distances to the camera, reducing large imaging-scale variations compared with more irregular geometries, although practical imaging performance is still affected by lens distortion, refraction, and camera placement.

Finally, the spherical design is consistent with common ergonomic objects used for sustained hand contact, including computer mice, ball-shaped doorknobs, and therapy grip balls, which often adopt rounded surfaces that match the natural wrapping posture of the hand. This design convergence provides additional ergonomic support for using a curved interface.

Together, the hand-motion statistics and these system-level considerations show that a spherical geometry is well suited for FusionTouch. The sphere provides broad hand-workspace coverage, ergonomic support, repeated-sector manufacturability, and a compact optical configuration, making it a balanced design choice for the proposed glove-free multimodal haptic interface.

Supplementary Text 2 Quantitative Assessment of Device Wearing Comfort

Device wearing comfort was evaluated using an 11-item questionnaire based on a 7-point Likert scale (*1*), where 1 indicated "strongly disagree", 4 indicated "neutral", and 7 indicated "strongly agree". The questionnaire covered five aspects: comfort and fit, movement flexibility, material feel and breathability, wearing-induced fatigue, and wearing convenience. The evaluated items included finger pressure (C1), hand-size suitability (C2), local discomfort (C3), natural finger bending (FL1), tactile feedback perception (FL2), skin comfort (M1), breathability (M2), sweating or friction (M3), hand fatigue (FT1), finger soreness (FT2), and ease of wearing/removal (W1).

For each participant, the raw response to item k was denoted as S_k^{raw}. Positive items were scored directly, whereas negative items, including finger pressure, local discomfort, hand fatigue, and finger soreness, were reverse-scored:

$$S_k = \begin{cases} S_k^{\mathrm{raw}}, & \text{positive item,} \\ 8 - S_k^{\mathrm{raw}}, & \text{negative item.} \end{cases} \tag{S1}$$

After correction, all item scores ranged from 1 to 7, with higher values indicating better wearing comfort.

The score of each comfort dimension was calculated as

$$D_d = \frac{1}{n_d} \sum_{k \in d} S_k, \tag{S2}$$

where D_d is the score of dimension d and n_d is the number of items in that dimension. The overall comfort score was calculated as the mean of all corrected item scores:

$$C = \frac{1}{11} \sum_{k=1}^{11} S_k. \tag{S3}$$

For intuitive comparison, the overall score was normalized to a 0–100 comfort index:

$$CI = \frac{C - 1}{6} \times 100. \tag{S4}$$

After converting the corrected 7-point Likert scores to a percentage-based comfort satisfaction score, FusionTouch achieved an average satisfaction of 84.82%, whereas the teleoperation glove achieved 40.54%. This comparison indicates that the spherical glove-free interface provides substantially higher perceived wearing comfort than the glove-based teleoperation interface.

Supplementary Text 3 Stiffness Enhancement via Outer-Layer Inflation Confinement: A Numerical Study

To examine the influence of the protective outer layer of FusionTouch on the contact stiffness of the structure, the indentation responses of systems with and without the outer layer are calculated and compared through Finite Element Method (FEM) simulations.

Here, we simplified the multi-layer optoelectronic skin to a double-layer spherical shell system. The outer thin shell represents the protective layer with a thickness of 0.2 mm and a radius of 70.5 mm, while the inner thick shell represents the thermoplastic rubber (TPR) support layer with a thickness of 2 mm and a radius of 69 mm. Thus, an interval of 0.4 mm is arranged between the outer shell and the inner shell. A circular opening with a diameter of 100 mm is designed at the bottom of these shells for reproducing the connection of the basement, which is set as a clamped boundary in the simulation.

Considering that the protective layer is made of nylon fabric and experiences only small strain, we adopted an isotropic linear elastic material model with a Young's modulus of 1 GPa and a Poisson's ratio of 0.49 for the outer shell. As for the inner shell, a hyperelastic material of the Mooney-Rivlin model is applied, of which the material parameters are determined by the tensile test data in Fig. **S5**(A).

To simulate the interface contact, a rigid indenter with a diameter of 30 mm and a contact radius of 25 mm is modeled to poke the sphere at the top. Surface-to-surface interaction is selected for all the contacting surfaces, which is defined as "Hard" in the normal direction. Before loading with the indenter, the inner layer shell is pressurized internally with an amount of 10 mbar, which is kept constant throughout the indentation process.

For the double-sphere model, it includes both the inner thick shell and the outer thin shell. For the single-sphere model, only the inner thick shell is contained. The simulation results of the force-displacement curves are presented in Fig. **S5**(B) along with the deformation images. Compared to the single-sphere model, the double-sphere model exhibits a much higher stiffness under the same indentation and pressure conditions. Viewing the deformation images of the FEM simulation, the involvement of the outer layer can effectively confine the pressurized inflation of the inner layer, thereby resulting in a high structural stiffness.

Supplementary Text 4 Near-infrared Image Processing

Although the near-infrared image contains implicit deformation cues through marker displacement, the raw image is not directly compatible with a graph neural network, which requires an explicit set of nodes and their spatial relationships. We therefore design an image preprocessing pipeline to extract reliable marker-level primitives from the near-infrared images before constructing the graph representation.

First, gamma correction is applied to normalize and enhance the image appearance (*2, 3*). In this step, the gamma value is estimated by matching the foreground brightness distribution of the camera images to that of a reference image set. The corrected images show higher contrast between the markers and the background, making the subsequent marker segmentation step more stable.

Second, a U-Net-based segmentation network (*4*) is used to generate binary marker masks from the enhanced images. Each input image is resized to the training resolution, normalized, and passed through the trained U-Net model. The output probability map is converted into a binary mask using a fixed threshold, where the foreground pixels correspond to the detected marker regions. The predicted masks are saved with the original filenames for subsequent processing.

Finally, morphological post-processing is performed to remove noise and separate markers that are incorrectly connected in the binary masks. Connected components with normal area and shape are preserved unchanged, while overly large or elongated components are treated as merged marker blobs. For these merged regions, erosion and dilation are applied to break thin bridges between neighboring markers, and small residual components are removed. This step produces cleaner marker masks while preserving the original shape of isolated markers.

The cleaned binary marker masks are then used to select valid marker regions and construct the marker graph, where each detected marker is treated as a graph node and the spatial relationships among neighboring markers define the graph structure.

Supplementary Text 5 Surface Deformation Reconstruction Network

We propose a graph-based coarse-to-fine 3D reconstruction network to recover the surface deformation of the tactile dome from a single marker image (*5*). The detected markers are first converted into a graph representation. Stage 1 deforms a 256-point tactile template to predict a coarse point cloud, which captures the global contact location and deformation profile. Stage 2 then uses the dense template mesh as the reconstruction topology, interpolates information from the coarse point cloud to the dense vertices, and predicts residual refinements to obtain the final dense surface reconstruction. To avoid using oracle contact information during inference, the contact area used by the deformation decoder is predicted from the tactile observation, whereas annotation-derived contact labels are used only for training supervision.

Problem formulation and marker graph. Let $\mathcal{M}_0 = (V_0, F_0)$ denote the undeformed template mesh of the tactile dome, where $V_0 \in \mathbb{R}^{N_v \times 3}$ contains the template vertex positions and F_0 denotes the fixed mesh connectivity. For each tactile observation, the measured surface geometry is denoted as $V^* \in \mathbb{R}^{N_v \times 3}$ after template-based normalization and nearest-neighbor matching. The network predicts a deformed mesh $\hat{V} \in \mathbb{R}^{N_v \times 3}$ that approximates V^* while preserving the template topology.

Each tactile image is represented as a marker graph rather than raw pixels (*6, 7*). For each detected marker, we use an 8-dimensional descriptor,

$$\mathbf{f}_i = \left[x_i^{\mathrm{rel}}, y_i^{\mathrm{rel}}, \tilde{a}_i, r_i, \eta_i, \sin 2\theta_i, \cos 2\theta_i, b_i\right]^{\mathsf{T}} \in \mathbb{R}^8, \tag{S5}$$

where x_i^{rel} and y_i^{rel} are the normalized marker centroid coordinates, $\tilde{a}_i$ is the normalized marker area, r_i and η_i describe the marker shape, θ_i is the fitted marker orientation, and b_i indicates whether the marker is close to the image boundary. The doubled-angle representation $\sin 2\theta_i$ and $\cos 2\theta_i$ removes the π-period ambiguity of ellipse orientation. Neighboring markers are connected using Delaunay triangulation, and each edge stores normalized relative geometric information between adjacent markers. The resulting graph captures both local marker deformation cues and the global spatial layout of the marker field.

Contact-mask generation and inference. During supervised training, contact masks are generated from the measured surface deformation and used as contact-aware training gates and region

labels. For the coarse template point set $P_0 = \{p_i^0\}_{i=1}^{N_p}$, where $N_p = 256$, the measured surface point cloud is nearest-neighbor matched to the template to obtain $P^* = \{p_i^*\}_{i=1}^{N_p}$. The coarse contact mask is defined as

$$m_i^c = \mathbb{I}\left(\left\|p_i^* - p_i^0\right\|_2 > \tau_c\right), \tag{S6}$$

where τ_c is the deformation threshold and $\mathbb{I}(\cdot)$ denotes the indicator function. Similarly, the dense contact mask m_j^f is obtained by thresholding $\|v_j^* - v_j^0\|_2$ for each dense mesh vertex.

To avoid using deformation-derived masks during testing, we additionally train an auxiliary contact-mask estimator that predicts contact regions directly from the tactile marker observation. The estimator takes the marker graph as input, extracts a global tactile latent code and marker-node features using a GNN encoder, and predicts a contact probability $\hat{s}_i^c \in [0, 1]$ for each coarse template point by aggregating nearby marker features around that point. The estimator is supervised using the coarse contact masks in Eq. (S6), with positive contact points upweighted to account for the imbalance between contact and non-contact regions.

During held-out evaluation and inference, measured deformation annotations are unavailable and are not provided to the reconstruction network. The annotation-derived masks are therefore replaced by the predicted contact probabilities $\hat{s}_i^c$, or by their thresholded binary masks,

$$\hat{m}_i^c = \mathbb{I}\left(\hat{s}_i^c > \tau_m\right), \tag{S7}$$

where τ_m is the contact-probability threshold. This separation ensures that ground-truth deformation is used only to construct supervised training labels, whereas test-time reconstruction relies only on the tactile marker observation.

Coarse deformation prediction. A graph encoder aggregates marker-node features and edge attributes into a latent code z describing the tactile observation. Conditioned on z, the coarse decoder predicts a displacement field $\Delta P = \{\Delta p_i\}_{i=1}^{N_p}$ on the 256-point template set P_0. The predicted coarse point cloud is obtained by applying a contact-aware gate,

$$\hat{p}_i = p_i^0 + g_i^c \Delta p_i, \tag{S8}$$

where $g_i^c = m_i^c$ during supervised training and $g_i^c = \hat{s}_i^c$ or $g_i^c = \hat{m}_i^c$ during held-out evaluation and inference. This gate suppresses displacement in non-contact regions and focuses deformation prediction on the contact area estimated from the marker observation at test time.

The coarse stage is supervised using a contact-region offset reconstruction loss,

$$\mathcal{L}_{\text{coarse}} = \frac{1}{3N_p^+} \sum_{i=1}^{N_p} m_i^c \left\| (\hat{p}_i - p_i^0) - (p_i^* - p_i^0) \right\|_1 , \tag{S9}$$

where $N_p^+ = \max(\sum_{i=1}^{N_p} m_i^c, 1)$ is the number of contact points on the coarse template, with a lower bound used to avoid division by zero. The annotation-derived mask in Eq. (S6) is used here only because the training labels provide the measured deformation field.

Fine mesh refinement. The fine stage upsamples the Stage-1 coarse point cloud into a dense mesh reconstruction. The dense template mesh provides the fixed output topology and the query vertices for interpolation. Given the coarse point cloud $\hat{P}$ and the latent code z, the fine module first extracts multi-scale geometric features from $\hat{P}$ using EdgeConv layers (*8*). For each dense template vertex v_j^0, the coarse-guided base position and coarse feature are interpolated from the k nearest coarse points using inverse-distance weighting,

$$\tilde{v}_j = \sum_{i \in \mathcal{N}_k(j)} w_{ji} \hat{p}_i, \qquad \tilde{h}_j = \sum_{i \in \mathcal{N}_k(j)} w_{ji} h_i, \tag{S10}$$

where h_i is the coarse feature at point $\hat{p}_i$, $\mathcal{N}_k(j)$ denotes the k nearest coarse points associated with vertex v_j^0, and w_{ji} are normalized inverse-distance weights.

The fine-stage contact gate is obtained from the available coarse contact gate. During training, the dense gate is the annotation-derived dense mask m_j^f. During held-out evaluation and inference, the predicted coarse contact mask is propagated to the dense mesh vertices by nearest-neighbor assignment or interpolation, yielding $\hat{m}_j^f$. We denote the resulting fine-stage gate as g_j^f. The fine decoder predicts a bounded residual displacement for each dense vertex from the interpolated feature, the latent code, and the coarse-guided base position,

$$\Delta v_j = s \tanh\left(\frac{\text{MLP}_\psi\left([\tilde{h}_j;\ z;\ \tilde{v}_j]\right)}{s} \right), \tag{S11}$$

where MLP_ψ denotes a multilayer perceptron and s limits the residual magnitude. The final dense vertex prediction is

$$\hat{v}_j = \tilde{v}_j + g_j^f \Delta v_j . \tag{S12}$$

Thus, Stage 2 refines the coarse geometry rather than generating a dense surface independently from the template. The dense template mesh defines the output topology, the coarse prediction provides the low-frequency deformation structure, and the residual decoder recovers local geometric details.

Fine-stage objective. The fine-stage loss combines region-wise reconstruction losses with mesh regularization,

$$\mathcal{L}_{\text{fine}} = \lambda_c \mathcal{L}_{\text{contact}} + \lambda_n \mathcal{L}_{\text{nc}} + \lambda_\ell \mathcal{L}_{\text{lap}} + \lambda_e \mathcal{L}_{\text{edge}}. \tag{S13}$$

The contact reconstruction loss is

$$\mathcal{L}_{\text{contact}} = \frac{1}{3N_v^+} \sum_{j=1}^{N_v} m_j^f \left\| \hat{v}_j - v_j^* \right\|_1, \tag{S14}$$

where $N_v^+ = \max(\sum_{j=1}^{N_v} m_j^f, 1)$. The non-contact loss is

$$\mathcal{L}_{\text{nc}} = \frac{1}{3N_v^-} \sum_{j=1}^{N_v} \left(1 - m_j^f\right) \left\| \hat{v}_j - v_j^* \right\|_1, \tag{S15}$$

where $N_v^- = \max(\sum_{j=1}^{N_v} (1 - m_j^f), 1)$. The dense contact mask is used to separate contact and non-contact regions in the supervised loss. This region-wise supervision helps preserve the global dome geometry while allowing local deformation in the contact region.

The Laplacian smoothness loss regularizes local surface variation by penalizing differences between each vertex and the average of its neighboring vertices,

$$\mathcal{L}_{\text{lap}} = \frac{1}{3N_v} \sum_{j=1}^{N_v} \left\| \hat{v}_j - \frac{1}{|\mathcal{N}(j)|} \sum_{k \in \mathcal{N}(j)} \hat{v}_k \right\|_1, \tag{S16}$$

where $\mathcal{N}(j)$ denotes the one-ring neighboring vertices of vertex j. This term suppresses isolated noisy residuals and encourages locally smooth dense surface reconstruction.

The edge-length regularization loss preserves the local mesh structure by penalizing changes in edge lengths relative to the undeformed template mesh,

$$\mathcal{L}_{\text{edge}} = \frac{1}{|\mathcal{E}|} \sum_{(j,k) \in \mathcal{E}} \left| \left\| \hat{v}_j - \hat{v}_k \right\|_2 - \left\| v_j^0 - v_k^0 \right\|_2 \right|, \tag{S17}$$

where $\mathcal{E}$ is the edge set of the template mesh. This term discourages excessive stretching or shrinking of local mesh edges, thereby reducing mesh distortion during fine residual refinement.

Training. The surface reconstruction network is trained in two phases. In Phase 1, the marker-graph encoder and coarse decoder are trained using the contact-region reconstruction loss in Eq. (S9). The annotation-derived coarse contact masks are used as training gates and region labels

because the measured deformation field is available during supervised training. In Phase 2, the best Phase-1 checkpoint is loaded and the fine mesh refinement module is introduced. The fine module is optimized with Eq. (S13), while the coarse module can be fine-tuned with a reduced learning rate. The checkpoint with the lowest validation Chamfer distance is selected for held-out evaluation. During held-out evaluation and inference, the annotation-derived contact masks are not used; they are replaced by contact masks predicted from the tactile marker observation.

Supplementary Text 6 Contact Force Estimation Network

We propose a pressure-conditioned patch-graph network to estimate the three-axis contact force from a single tactile observation. The network uses marker-centered image patches to capture local deformation appearance and represents the detected markers as a graph to encode their spatial layout. The internal pressure is incorporated as a conditioning signal, enabling the model to account for changes in the mechanical response of the inflated tactile surface (*9, 10*). The network outputs the three-axis force vector

$$\hat{\mathbf{F}} = [\hat{F}_x, \hat{F}_y, \hat{F}_z]^\mathsf{T} \in \mathbb{R}^3. \tag{S18}$$

Problem formulation and marker graph. Given one tactile observation and the corresponding internal pressure p_{int}, the network predicts the contact-force vector $\hat{\mathbf{F}} \in \mathbb{R}^3$ that approximates the measured force $\mathbf{F} \in \mathbb{R}^3$. Each detected marker is represented by two types of information: a marker-centered image patch and the 8-dimensional geometric descriptor defined in Eq. (S5). The image patch provides local appearance cues around the marker, while the geometric descriptor encodes marker position, area, shape, orientation, and boundary information.

The detected markers are connected using Delaunay triangulation to form a marker graph. Edge attributes describe the relative geometry between neighboring markers, including their normalized distance and relative displacement. This graph representation allows the network to aggregate both local marker deformation cues and the global deformation pattern of the marker field (*11*).

Patch and pressure-conditioned feature encoding. For each marker, a patch encoder extracts a local visual feature from the marker-centered image patch. To incorporate the inflation state of the tactile dome, the internal pressure p_{int} modulates intermediate visual features through FiLM (*12*),

$$\text{FiLM}(\mathbf{x}, p_{\text{int}}) = \gamma(p_{\text{int}}) \odot \text{BN}(\mathbf{x}) + \beta(p_{\text{int}}), \tag{S19}$$

where $\gamma(p_{\text{int}})$ and $\beta(p_{\text{int}})$ are pressure-dependent affine parameters, $\text{BN}(\cdot)$ denotes batch normalization, and $\odot$ denotes element-wise multiplication. This modulation allows the same visual deformation pattern to be interpreted differently under different internal pressures.

In addition, a pressure-aware scalar is appended to each node feature,

$$\phi_i = \left(\frac{p_{\text{int}}}{p_{\max}}\right) \tilde{a}_i, \tag{S20}$$

where $\tilde{a}_i$ is the normalized marker area and $p_{\max}$ is the maximum pressure used for normalization. This scalar provides a pressure-dependent cue relating marker appearance change to the current inflation state.

Graph-based force regression. The initial node feature is formed by concatenating the pressure-conditioned patch feature, the geometric descriptor in Eq. (S5), and the pressure-aware scalar in Eq. (S20). A graph encoder then propagates information over the marker graph, allowing each marker feature to be updated using both its local appearance and neighboring marker geometry. The updated node features are pooled into a global tactile representation, which is passed to a multilayer regression head to predict the three-axis contact force.

Training objective. The force estimation network is trained with a composite regression loss,

$$\mathcal{L}_{\text{force}} = \lambda_{\text{F,mae}}\mathcal{L}_{\text{F,MAE}} + \lambda_{\text{F,rmse}}\mathcal{L}_{\text{F,RMSE}}, \tag{S21}$$

where the force MAE loss is defined as

$$\mathcal{L}_{\text{F,MAE}} = \frac{1}{3B}\sum_{b=1}^{B}\left\|\hat{\mathbf{F}}^{(b)} - \mathbf{F}^{(b)}\right\|_1, \tag{S22}$$

and the force RMSE loss is defined as

$$\mathcal{L}_{\text{F,RMSE}} = \sqrt{\frac{1}{3B}\sum_{b=1}^{B}\left\|\hat{\mathbf{F}}^{(b)} - \mathbf{F}^{(b)}\right\|_2^2 + \epsilon_{\text{num}}}. \tag{S23}$$

Here, B is the batch size and ϵ_{num} is a small constant for numerical stability. The MAE term encourages stable component-wise force prediction, whereas the RMSE term increases the penalty on larger force errors. The checkpoint with the lowest validation loss is selected for held-out evaluation.

Supplementary Text 7 Hand Pose Estimation Network

We use an image-based regression network to estimate hand pose from tactile bubble-view images. The network predicts a 12-dimensional joint-angle vector from either the visible-light image, the near-infrared image, or their dual-modality combination. The three input settings use the same target representation and training objective, enabling a consistent comparison across sensing modalities.

Problem formulation. Given one tactile image or a pair of tactile images, the network predicts the joint-angle vector

$$\hat{\mathbf{q}} = [\hat{q}_1, \ldots, \hat{q}_{12}]^\mathsf{T} \in \mathbb{R}^{12}, \tag{S24}$$

where each entry corresponds to a selected glove joint angle in degrees. The ground-truth joint-angle vector is denoted as $\mathbf{q} \in \mathbb{R}^{12}$. The predicted dimensions correspond to the selected glove joints r_0, r_1, r_2, r_4, r_5, r_7, r_8, r_9, r_{12}, r_{13}, r_{16}, and r_{17}.

Single- and dual-modality encoders. For single-modality estimation, the network takes either the visible-light image or the near-infrared image as input. An image encoder extracts a latent feature from the selected modality, and a regression head maps this feature to the 12-dimensional joint-angle vector. The visible-light and near-infrared single-modality models use the same architecture and differ only in the input modality.

For dual-modality estimation, the visible-light and near-infrared images are processed by two independent image encoders. Their latent features are concatenated and passed to a shared regression head for joint-angle prediction. This late-fusion design preserves modality-specific features before combining them for pose estimation.

Training objective. The hand-pose network is trained with a composite regression loss,

$$\mathcal{L}_{\text{pose}} = \lambda_{\text{q,mae}}\mathcal{L}_{\text{q,MAE}} + \lambda_{\text{q,rmse}}\mathcal{L}_{\text{q,RMSE}}, \tag{S25}$$

where the pose MAE loss is defined as

$$\mathcal{L}_{\text{q,MAE}} = \frac{1}{12B}\sum_{b=1}^{B}\left\|\hat{\mathbf{q}}^{(b)} - \mathbf{q}^{(b)}\right\|_1, \tag{S26}$$

and the pose RMSE loss is defined as

$$\mathcal{L}_{\mathrm{q,RMSE}} = \sqrt{\frac{1}{12B}\sum_{b=1}^{B}\left\|\hat{\mathbf{q}}^{(b)} - \mathbf{q}^{(b)}\right\|_2^2 + \epsilon_{\mathrm{num}}}. \tag{S27}$$

Here, B is the batch size and ϵ_{num} is a small constant for numerical stability. The MAE term encourages stable average prediction accuracy, whereas the RMSE term increases the penalty on larger angular errors. The same preprocessing, loss function, and model-selection protocol are used for all three input settings.

Supplementary Text 8 Hand-Motion Data Acquisition Tasks

To systematically evaluate the natural hand-motion workspace, we designed 20 representative daily bimanual manipulation tasks and used Manus Quantum Metagloves to collect hand-joint motion data during task execution. These tasks covered common daily operation patterns, including rotation, grasping, placement, pressing, insertion, writing, folding, pouring, pinching, and cable handling, with the aim of reproducing natural hand-motion states in realistic use scenarios.

Specifically, screw operation and screwing tasks were used to simulate rotational tool operations and to capture finger wrapping, thumb opposition, and wrist-assisted motion. Object sorting, block stacking, and placing tasks were used to simulate hand postures during grasping, transfer, alignment, and release. Paper cutting, stapling papers, writing, and erasing tasks recorded fine paper-based manipulation, continuous trajectory motion, and repetitive pressing or rubbing actions. Pouring beans and scooping rice captured dynamic gestures during container grasping, tilting, transfer, and pouring. Folding was used to simulate bimanual deformation of flexible materials. Plug insertion, USB insertion, wrench insertion, and battery replacement recorded precise alignment, insertion, assembly, and local adjustment motions. Chopstick use captured fingertip pinching and multi-finger coordination, whereas cable tightening and tangling wire simulated complex hand postures during flexible-cable handling, rotation, pulling, and winding.

To improve the naturalness and randomness of data acquisition, only the target of each task was specified during the experiment, while the exact motion trajectory, grasping strategy, operation speed, and finger force were not constrained. Participants were allowed to complete each task according to their own habitual operation patterns; therefore, inter-participant differences in finger flexion, contact position, motion path, and action rhythm were preserved in the dataset. In addition, task order, initial object placement, and starting hand posture were varied to reduce bias caused by a fixed experimental procedure and to make the collected data closer to the natural variability of daily hand operations.

To ensure the scientific validity of the samples, the 20 tasks were not designed around a single gesture type but instead covered multiple interaction objects, including rigid objects, flexible materials, granular materials, paper, tools, connectors, and cables. Different tasks corresponded to distinct dominant motion patterns, such as rotational manipulation, fine pinching, pressing,

insertion and alignment, trajectory writing, bimanual coordination, and flexible-object handling. Therefore, this task set provided broad coverage of the finger-joint activity range and hand-motion workspace commonly involved in daily activities.

To ensure fairness in data collection, all participants completed the same task set under the same experimental environment, device configuration, and acquisition procedure. Before formal recording, participants were allowed to become familiar with the data glove and task workflow to reduce abnormal movements caused by unfamiliarity with the equipment. During the experiment, participants' natural operation habits were not manually corrected, and they were not required to imitate a fixed motion template. This avoided introducing bias from a specific individual or a specific operation style. By combining unified task requirements with preserved individual motion differences, the collected data maintained cross-participant comparability while reflecting the diversity of hand motions in real daily operations.

Based on the motion data collected using the Manus data glove, the spatial distribution of finger nodes during natural manipulation was further analyzed. These data were used to calculate the hand-motion coverage ratio under different spherical diameters and to provide experimental support for determining the spherical-body diameter of FusionTouch.

Supplementary Figures

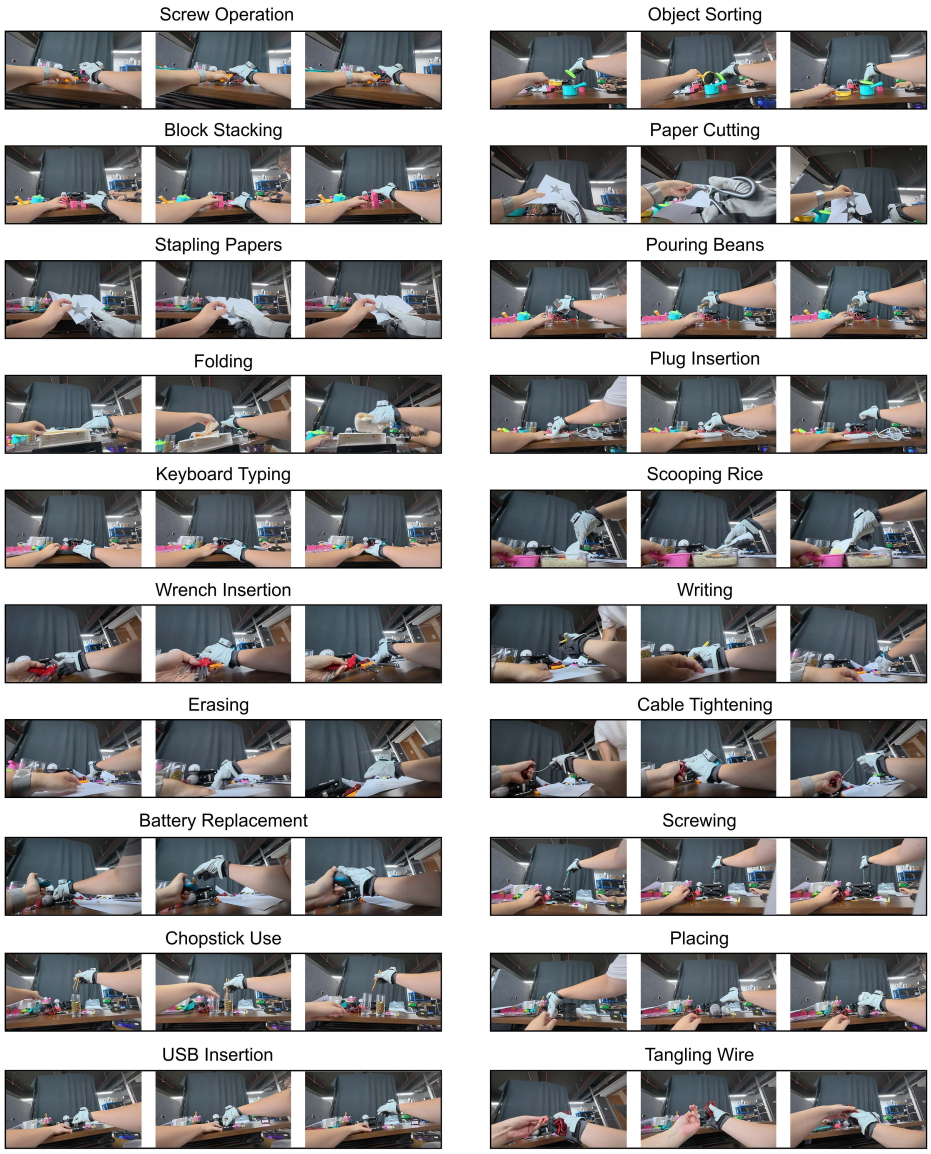


Fig. S1. Representative daily manipulation tasks used to collect Manus-glove hand-motion data for determining the FusionTouch spherical-body diameter.

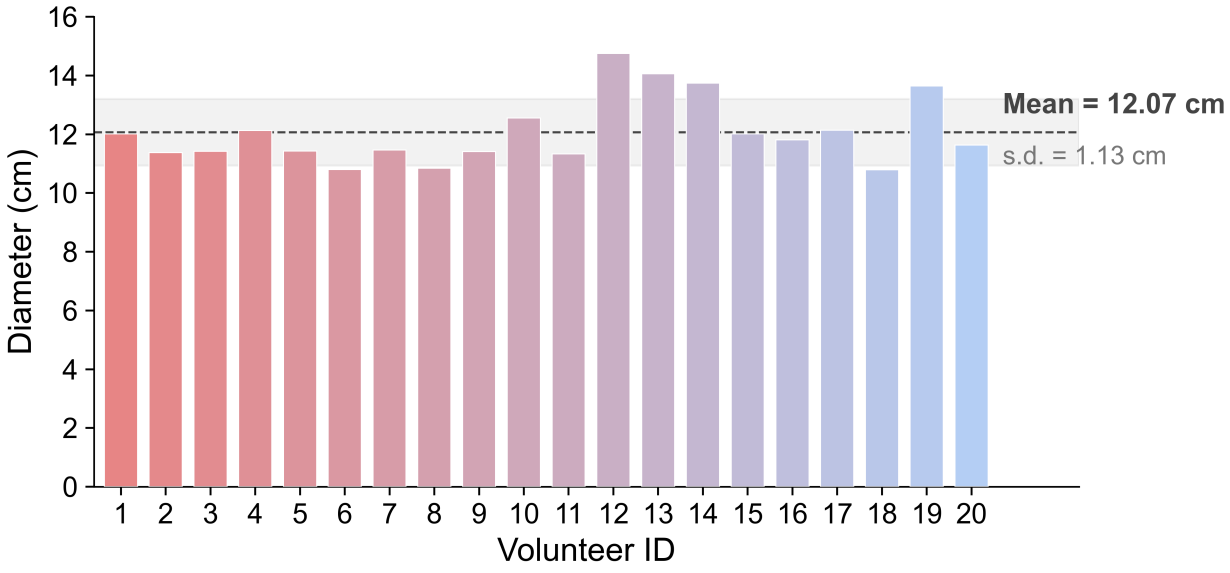


Fig. S2. **Maximum spherical-envelope diameter of naturally relaxed right hands across 20 volunteers.** Each bar represents the fitted sphere diameter estimated from the 3D hand model of one volunteer in the relaxed-hand state. The dashed horizontal line indicates the mean diameter across all volunteers, and the shaded gray region represents one standard deviation. The average diameter is 12.07 ± 1.13 cm.

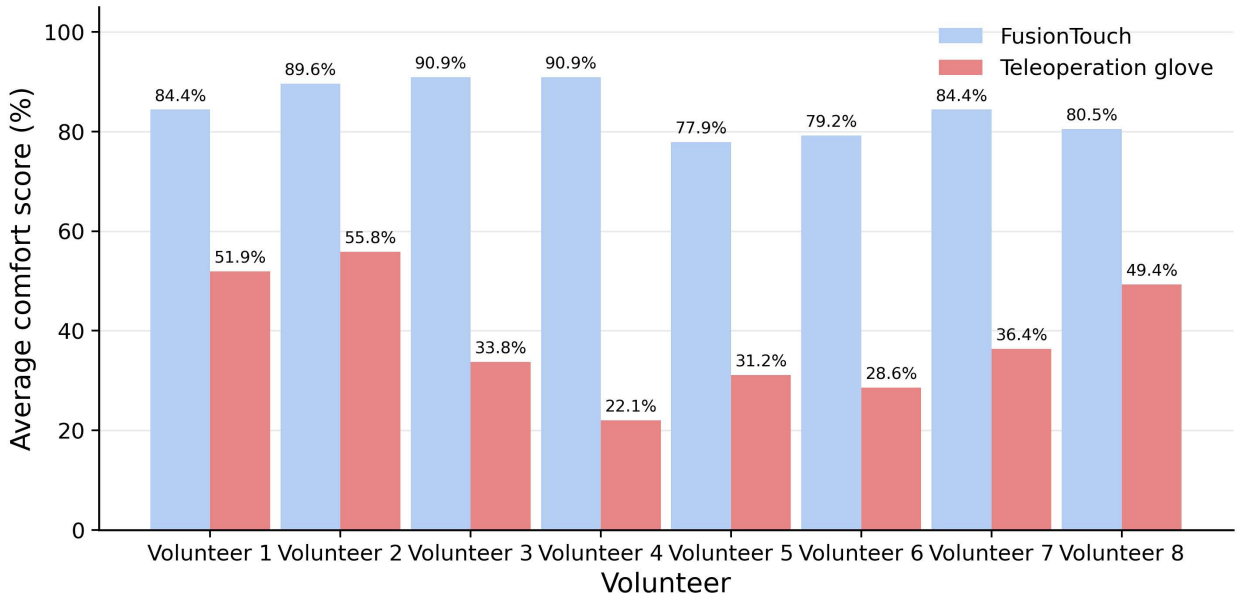


Fig. S3. **Participant-wise comfort scores for FusionTouch and a commercial data collection glove.** Each bar represents the normalized average comfort score of one volunteer, calculated from corrected 7-point Likert-scale responses and normalized to a 0–100 scale. Higher scores indicate better perceived wearing comfort.

Fig. S4. **Fourteen-electrode pyramidal configuration on the fabric cover.** The image shows the selected electrode layout used for the FusionTouch electrotactile array. Fourteen electrodes were distributed on the curved fabric cover in a pyramidal configuration, providing dense spatial coverage around the contact region while maintaining feasible wiring and mechanical integration. This layout was selected to improve stimulation localization and surface coverage on the soft curved interface, establishing the final channel count and spatial arrangement for the FusionTouch electrotactile module.

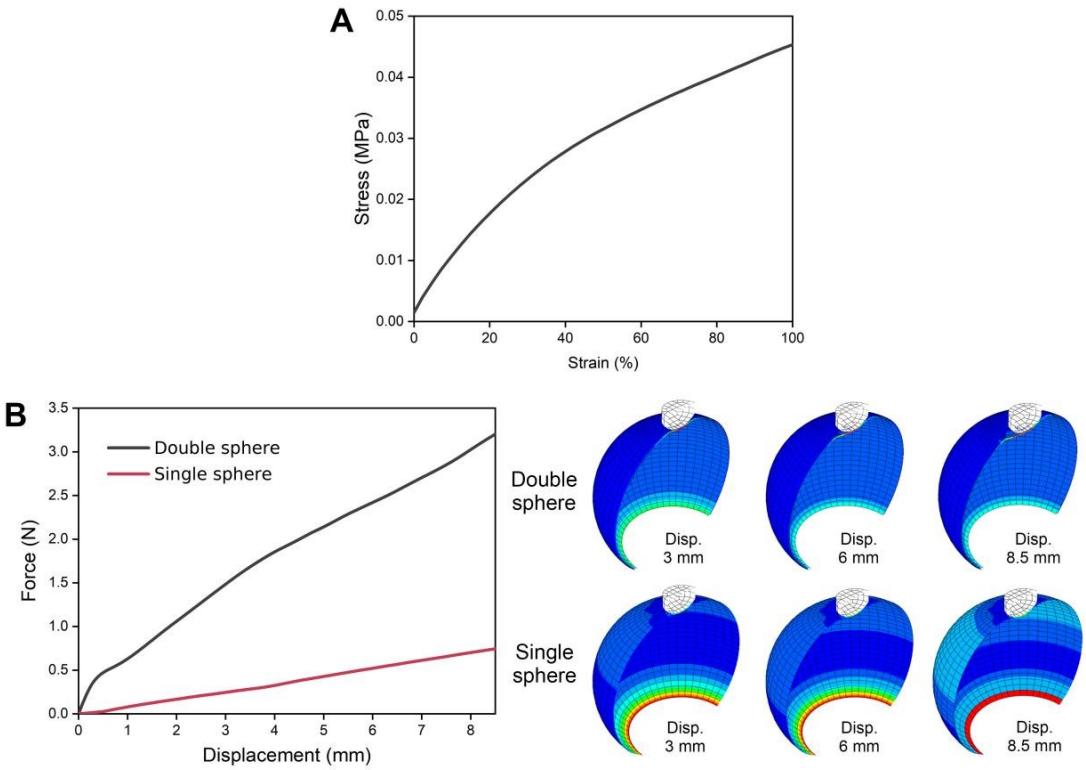
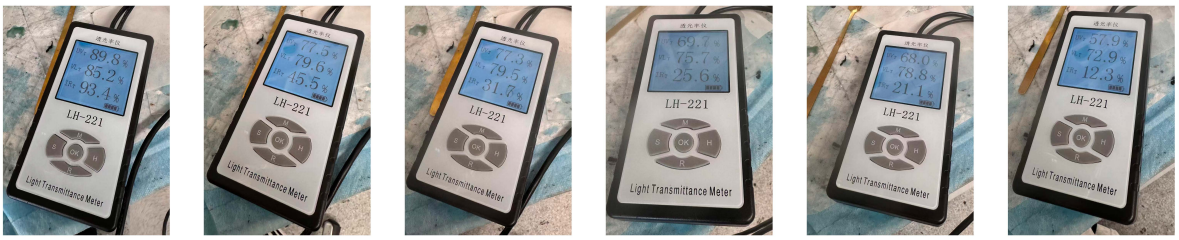


Fig. S5. **Comparison of the reaction force-indentation displacement curves given by FEM simulations of double-sphere and single-sphere models.** (A) Strain-stress curve of the TPR material given by the material tensile test. (B) Under the same internal pressure, the double-sphere model exhibits greater stiffness due to the geometric confinement of the outer layer, while the single-sphere model has a larger expansion volume but exhibits lower stiffness.



0% 0.5% 0.75% 1% 1.25% 1.5%

Fig. S6. **Physical transmittance measurements of infrared-absorbing coatings at different concentrations.** The images show the measurement process and light-transmittance-meter readings for coating samples with five concentrations: 0%, 0.5%, 0.75%, 1%, and 1.25%. For each sample, the ultraviolet, visible-light, and infrared transmittance values were recorded to supplement the spectral-selectivity results reported in the main text. As the coating concentration increased, the infrared transmittance decreased markedly, whereas the visible-light transmittance remained relatively high, indicating that the coating selectively attenuates the infrared channel while preserving visible-light observation.

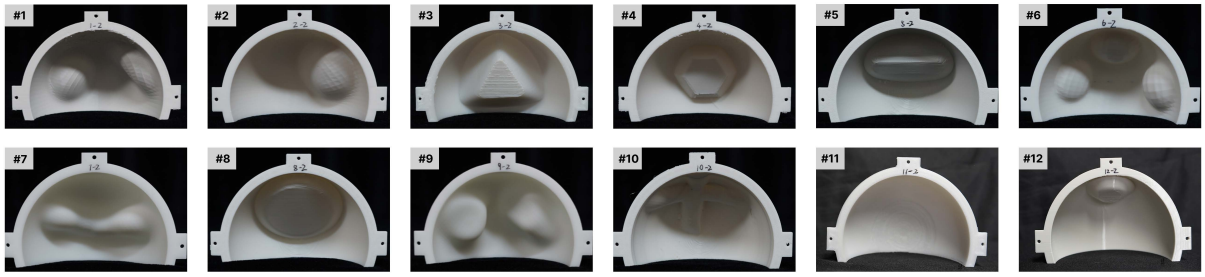


Fig. S7. **Calibration molds for 3D surface-reconstruction data collection.** The images show representative molds (#1–#12) with known inner-surface geometries used for FusionTouch 3D reconstruction calibration. Each mold contains different protrusions, depressions, and local curved features, producing controlled and repeatable deformations on the spherical soft surface. During data collection, one mold was placed over the outer surface of FusionTouch and rotated through 360° in 10° increments, yielding 36 orientations for each mold. The corresponding OBJ model of each mold was used to extract the inner-surface point cloud as the geometric ground truth for training and evaluating the 3D surface-reconstruction network.

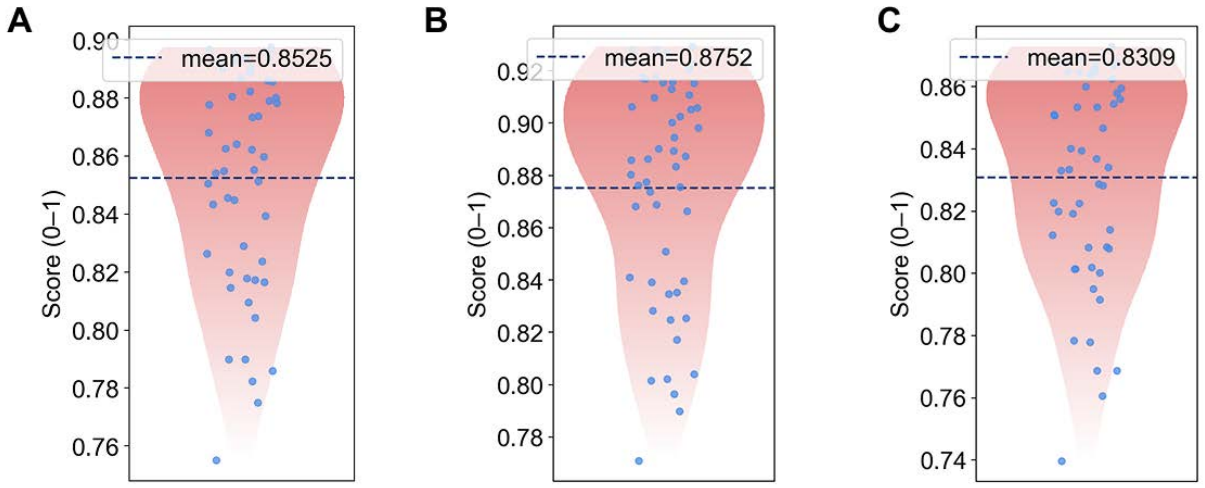


Fig. S8. **3D reconstruction performance across test samples.** (A) F-score distribution. (B) Precision distribution. (C) Recall distribution. Blue dots represent individual samples, shaded regions show the distributions, and dashed lines indicate the mean values.

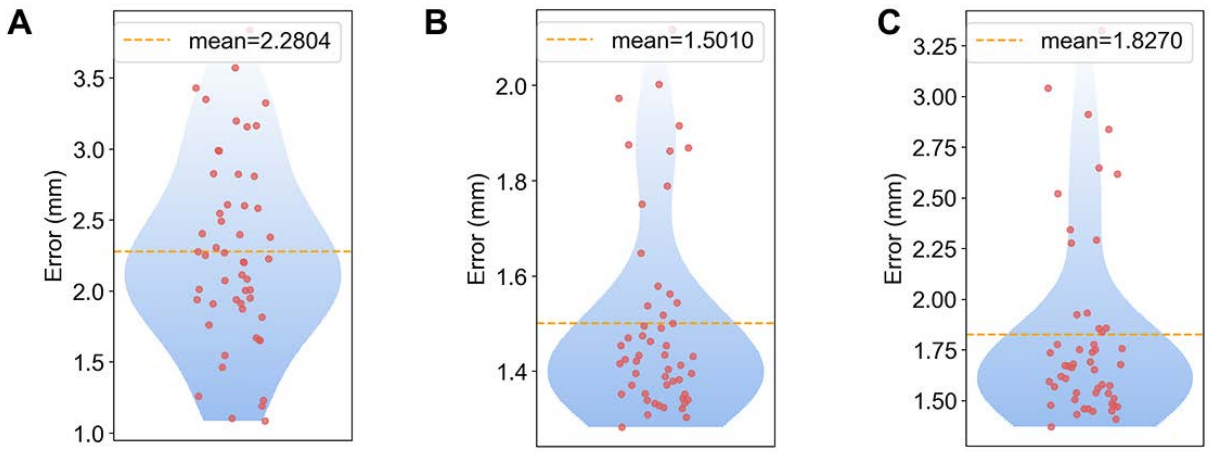


Fig. S9. **3D reconstruction error distributions across test samples.** (A) Coarse point cloud CD RMSE distribution. (B) Fine-grained point cloud point-to-surface (P2S) mean error distribution. (C) Fine-grained point cloud point-to-surface (P2S) RMSE error distribution. Red dots represent individual samples, shaded regions show the distributions, and dashed lines indicate the mean values. Errors are reported in millimeters, with lower values indicating better performance.

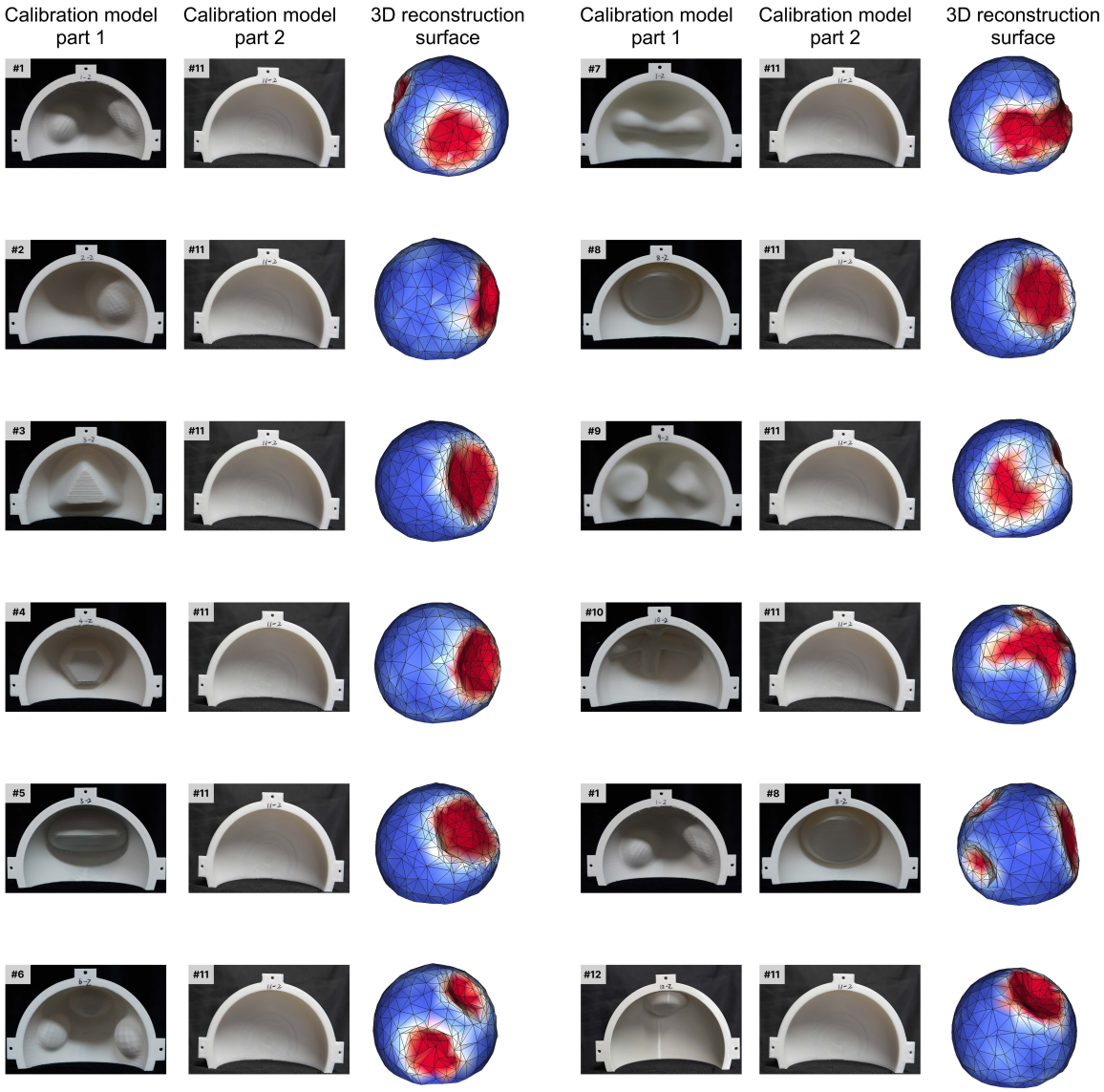


Fig. S10. **Representative surface-deformation reconstruction results under different calibration models.** For each numbered case, the upper image shows the physical calibration mold used to impose the contact deformation, and the lower image shows the corresponding reconstructed 3D surface of FusionTouch. The colored surface indicates the reconstructed deformation distribution on the spherical skin, where warmer colors correspond to larger contact-induced deformation.

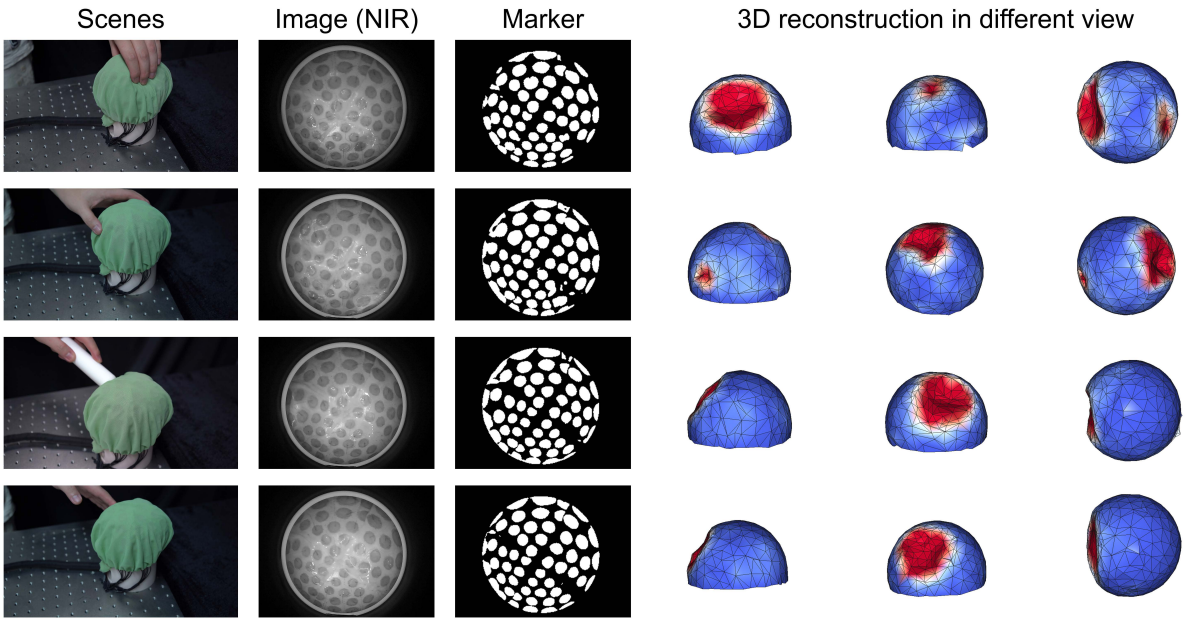


Fig. S11. **Qualitative visualization of 3D tactile reconstruction under different contact scenes.** Each row shows one representative interaction sample. From left to right, the columns show the external contact scene, the captured near-infrared tactile image, the extracted marker pattern, and the reconstructed 3D surface rendered from multiple viewpoints. The multi-view renderings visualize the spatial distribution of the reconstructed contact deformation, with red regions highlighting the main contact area on the sensor surface.

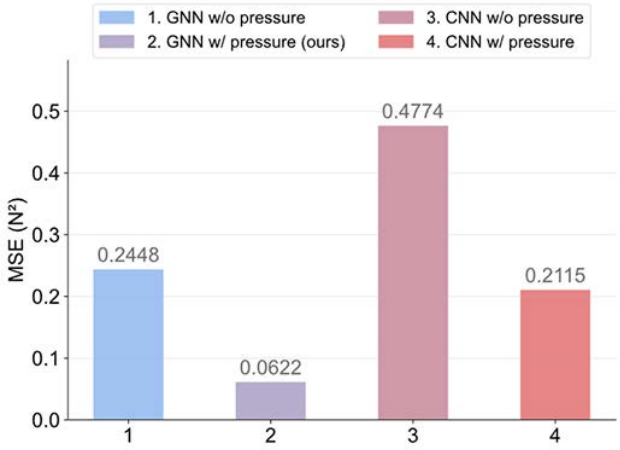


Fig. S12. **Mean squared error (MSE) comparison of the four methods for the total force magnitude.** The x-axis labels 1–4 correspond to GNN without pressure, GNN with pressure (ours), CNN without pressure, and CNN with pressure, respectively.

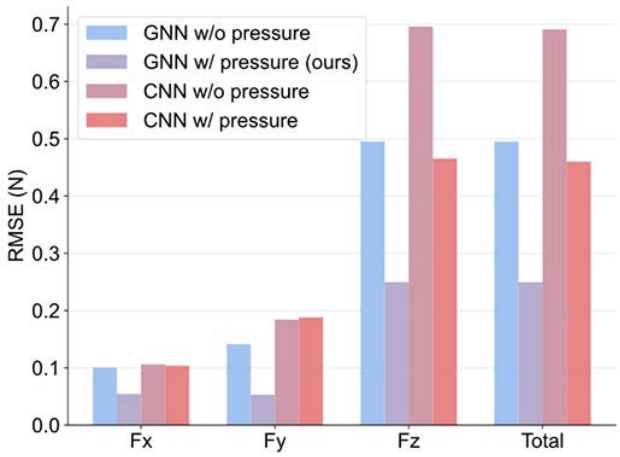


Fig. S13. Root mean square error (RMSE) comparison of the four force prediction methods across Fx, Fy, Fz, and the total force magnitude.

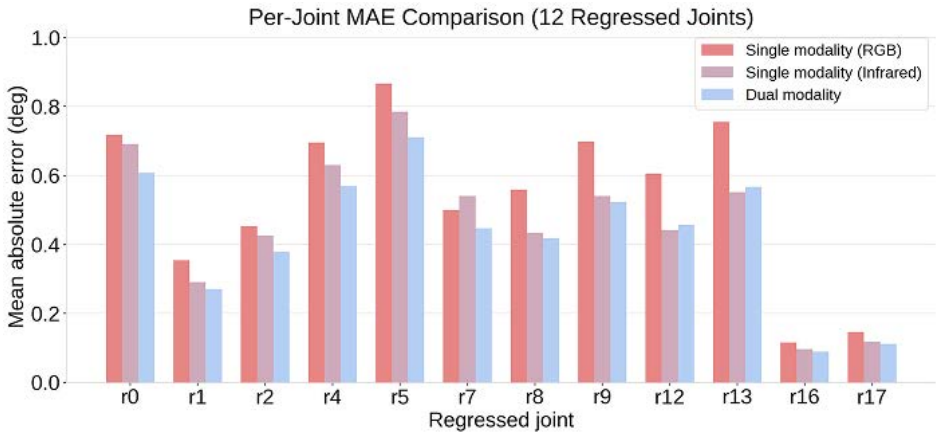


Fig. S14. **Per-joint hand-pose estimation errors across the 12 regressed joints.** Mean absolute error (MAE) is compared under single RGB, single infrared, and dual-modality inputs. Bars indicate the joint-level MAE for each input setting. Errors are reported in degrees, with lower values indicating better joint-angle prediction accuracy.

Fig. S15. **Radar comparison of the three camera-modality settings across overall pose-estimation metrics.** The dual-modality model shows the strongest overall performance, with lower error and higher threshold accuracy than the single-modality baselines.

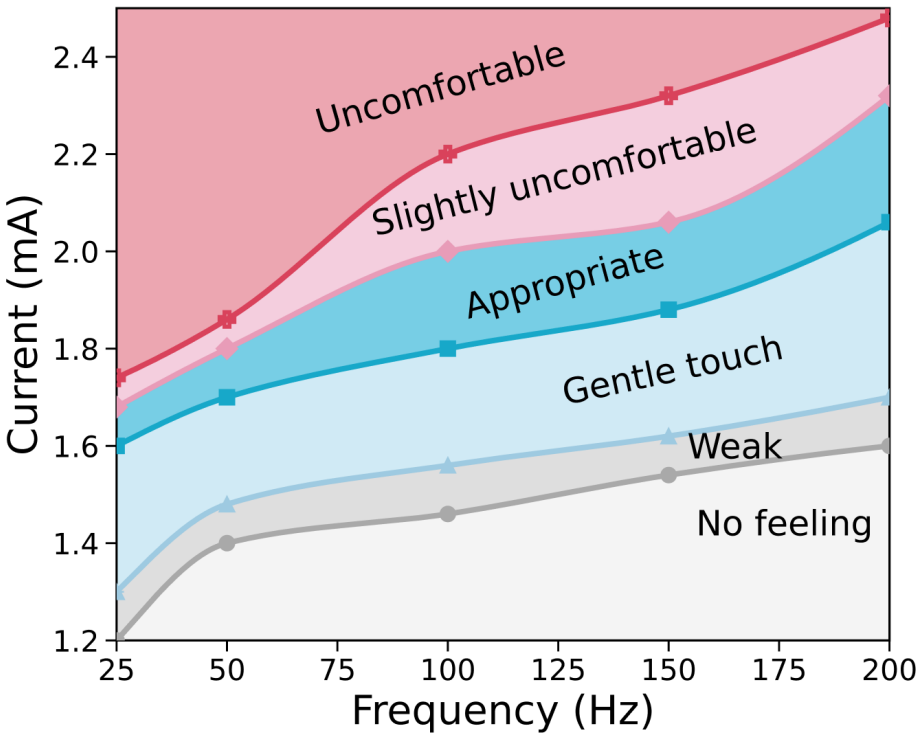


Fig. S16. **Relationship between stimulation frequency, current intensity, and subjective comfort levels.** The x-axis represents the stimulation frequency in Hz, and the y-axis represents the current intensity in mA. Based on five subjective rating thresholds measured at different frequencies, the stimulus space was divided into six perceptual regions: no feeling, weak, gentle touch, appropriate, slightly uncomfortable, and uncomfortable. The colored regions indicate the corresponding comfort zones as the current intensity varies with stimulation frequency.

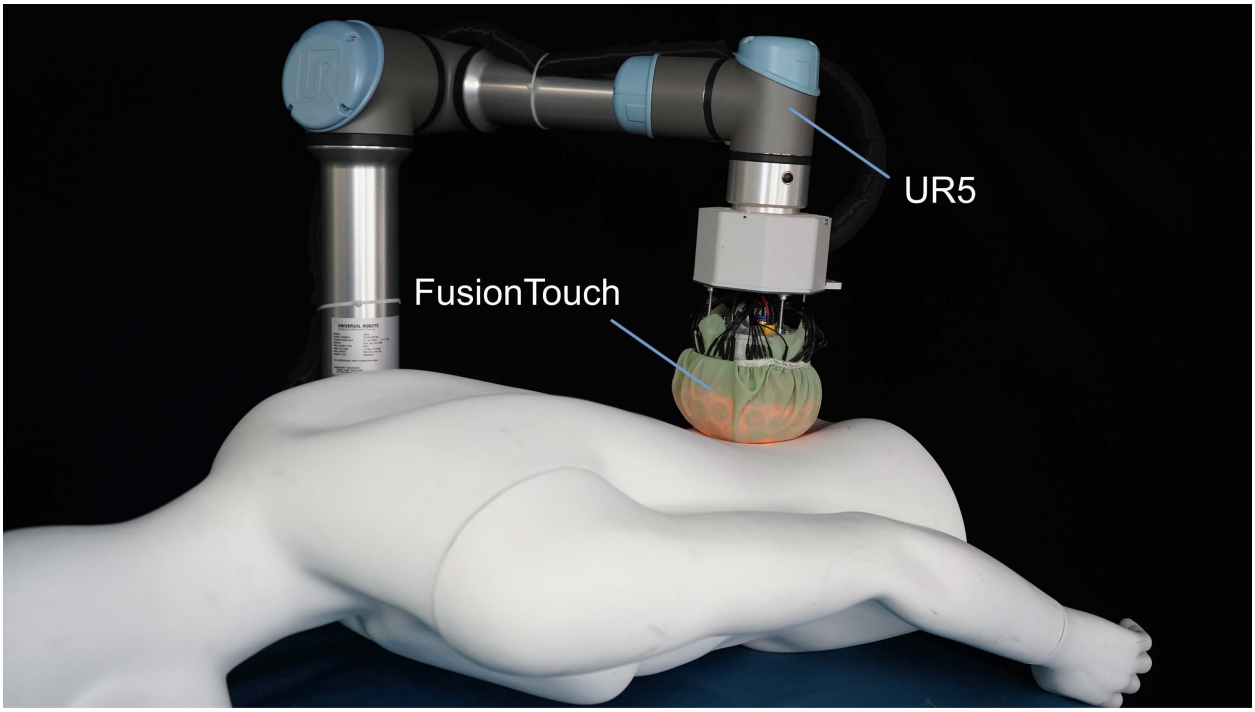


Fig. S17. **Robotic massage experiment setup.** The image shows the experimental platform used to evaluate FusionTouch during robotic massage interactions. A FusionTouch sensor was mounted as the end-effector of a UR5 robotic arm and brought into contact with the back region of a human-body phantom. The UR5 provided repeatable motion control, while FusionTouch measured the contact-induced deformation and near-infrared marker responses during massage. The robot end-effector pose was defined with respect to the robot base frame, and the contact direction was aligned approximately with the local surface normal of the phantom. During each trial, the robot followed a predefined massage trajectory while maintaining controlled contact with the surface, enabling synchronized acquisition of tactile images, deformation states, and robot motion parameters.

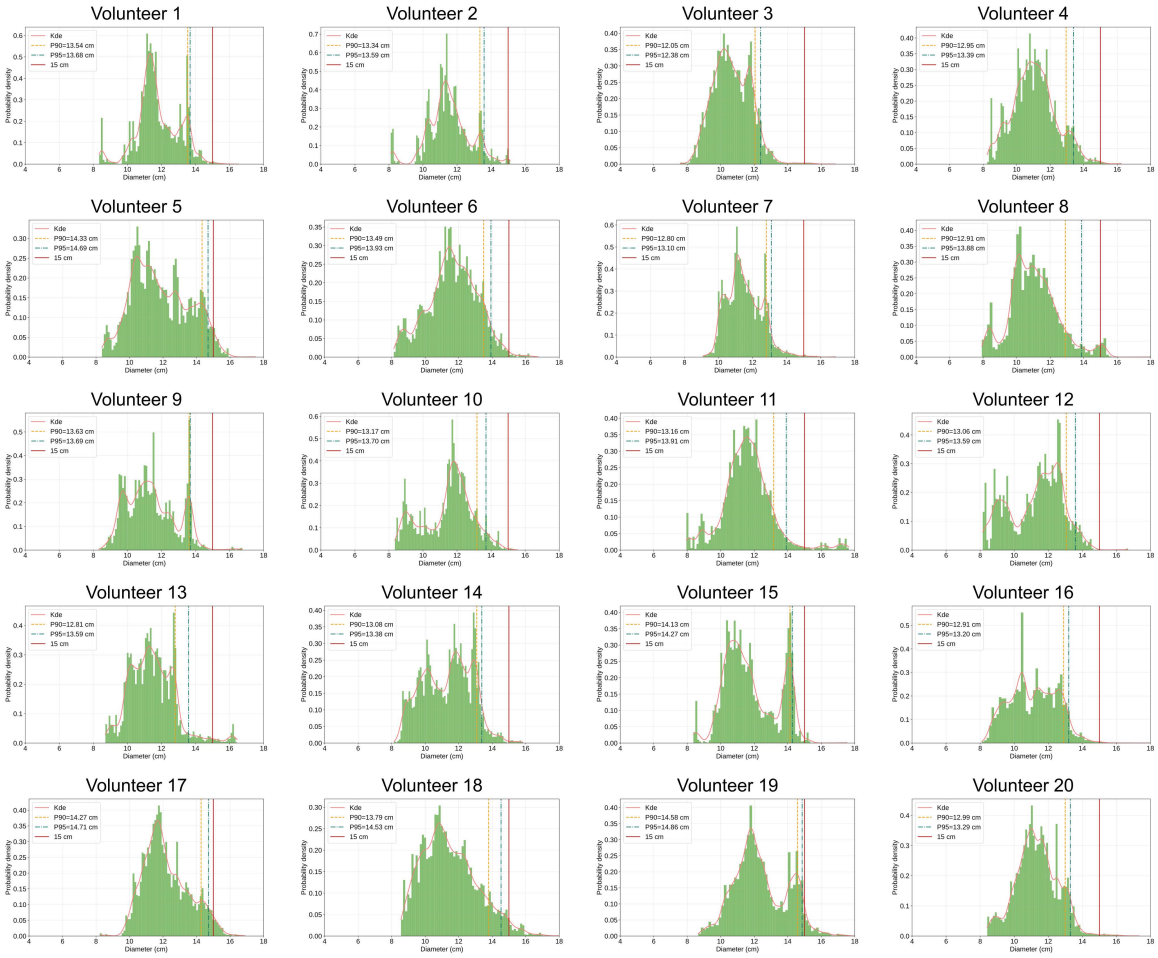


Fig. S18. **Hand-motion envelope diameter distributions across 20 volunteers.** Each subplot shows the probability-density distribution of frame-wise minimum-enclosing-sphere diameters computed from the 15 distal finger nodes of one volunteer. Vertical lines denote the P90 and P95 diameters and the 15 cm reference threshold, providing a statistical comparison between the recorded hand-motion envelope and the target spatial range.

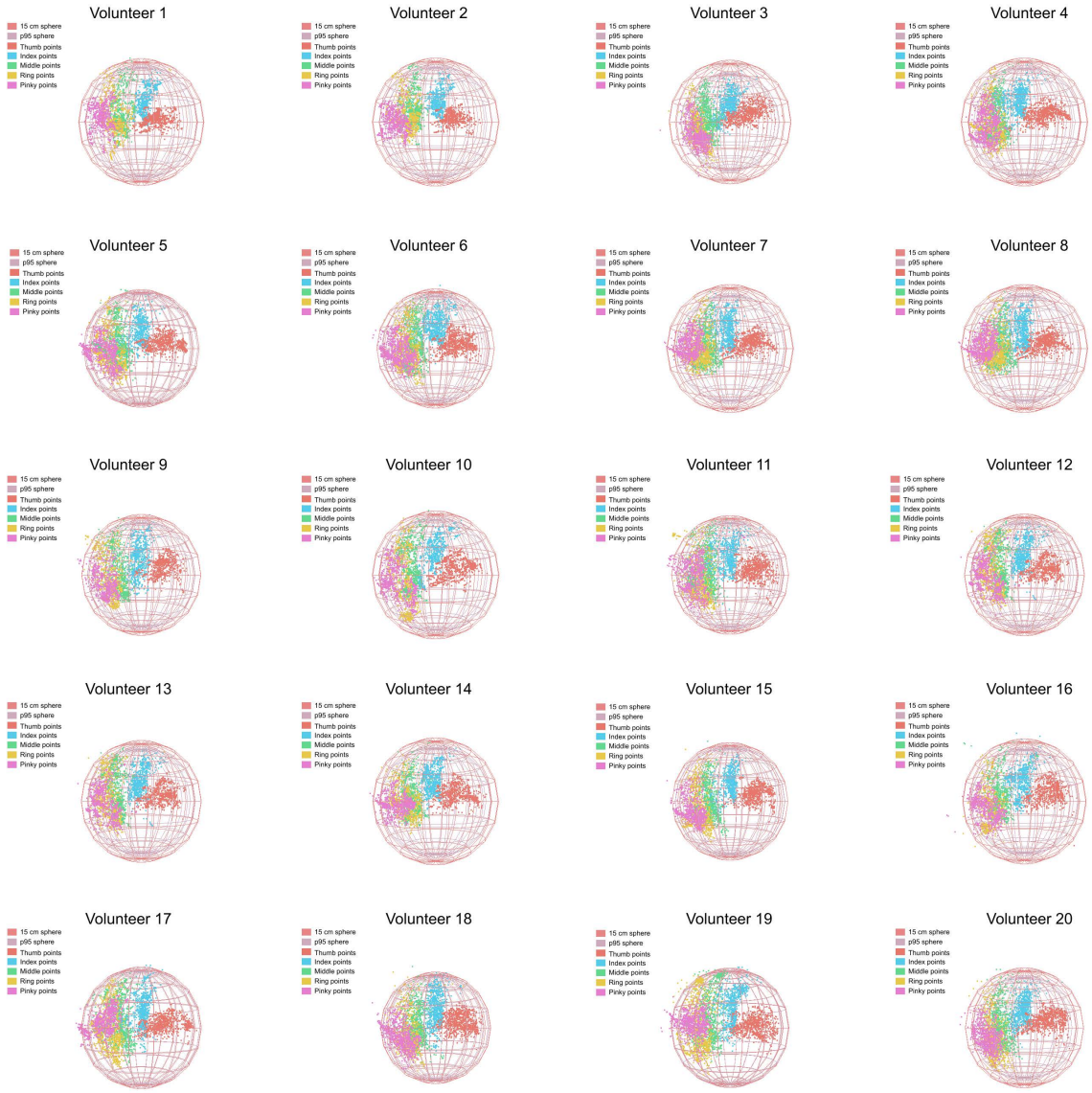


Fig. S19. **Hand-motion envelope distributions across 20 volunteers. Each subplot corresponds to one volunteer.** Colored scatter points show the aligned trajectories of distal finger nodes relative to the frame-wise minimum-enclosing-sphere center. The wireframe spheres denote the 15 cm reference sphere and the P95 motion-envelope sphere estimated from frame-wise 15-node statistics.

Fig. S20. **Single-view estimation of relaxed-hand 3D geometry.** For each volunteer (panels A–T), the naturally relaxed right hand is imaged using a single RGB camera and reconstructed using HAMER, which estimates a MANO-based 3D hand mesh from the input image. The reconstructed hand model is exported as an OBJ mesh for subsequent geometric analysis. Within each panel, the left image shows the original RGB capture, whereas the right image shows the estimated hand mesh rendered and overlaid on the corresponding hand image. The thin white line marks the boundary between the original capture and the overlaid reconstruction. Each capture contains a single relaxed right hand. The estimated hand models preserve the overall hand geometry and finger configuration in the natural relaxed state. The distal finger joints used to fit the node-enclosing sphere reported in Table S1 are derived from the reconstructed hand models.

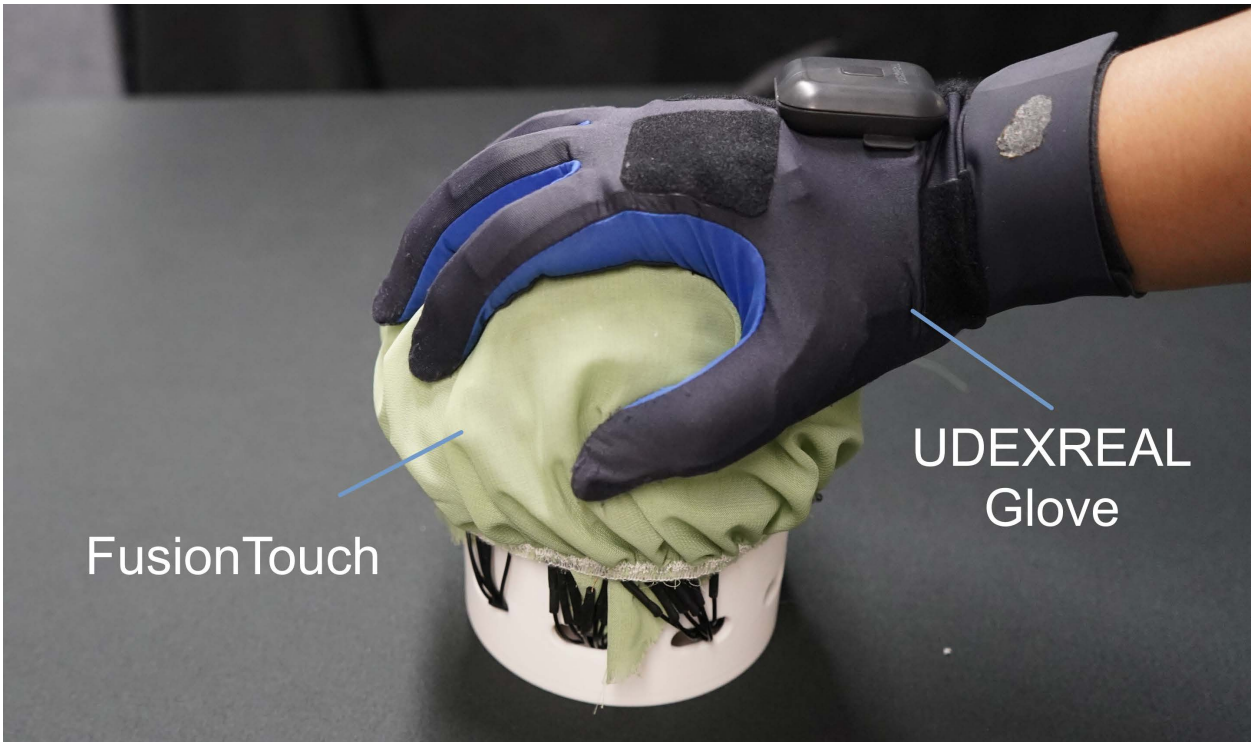


Fig. S21. **Hand-gesture data acquisition setup using the UDEXREAL data glove.** The image shows the experimental setup used to collect hand-joint motion data during interaction with FusionTouch. A participant wore a UDEXREAL data glove while grasping and deforming the FusionTouch sensing surface. The data glove recorded the joint states of the fingers during different hand gestures, while FusionTouch simultaneously captured contact-induced tactile responses from the soft sensing interface. This setup enabled synchronized acquisition of hand-gesture information and tactile deformation data, providing paired data for analyzing the relationship between finger joint configurations and tactile interaction patterns.

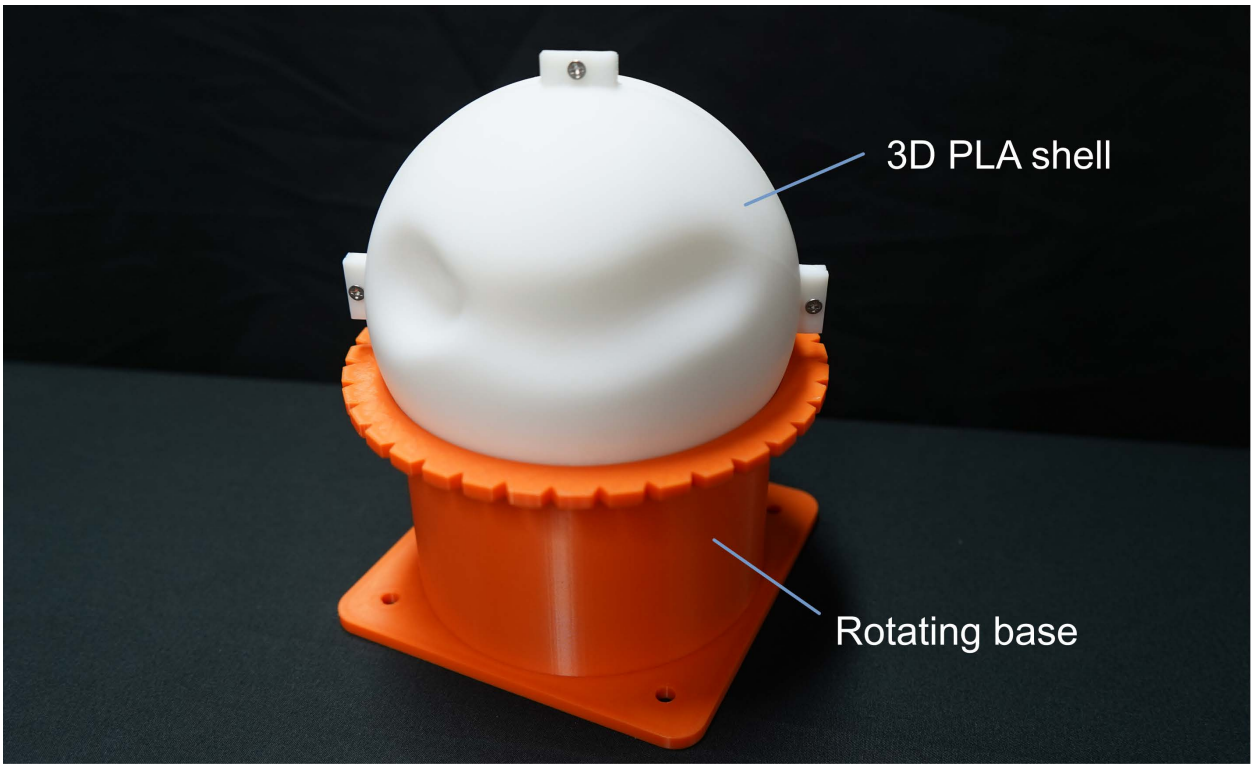


Fig. S22. **3D reconstruction dataset acquisition setup.** The image shows the calibration setup used to collect FusionTouch 3D surface-reconstruction data. A 3D-printed PLA shell with known surface geometry was mounted on an orange rotating base. The rotating base was divided into 36 indexed positions, with each position corresponding to a 10° rotation step. During data collection, the shell was rotated through 360° and one sample was acquired at each indexed position, yielding 36 orientations for each calibration shell. This setup enabled repeatable acquisition of near-infrared marker images under controlled deformation geometries for training and evaluating the 3D reconstruction model.

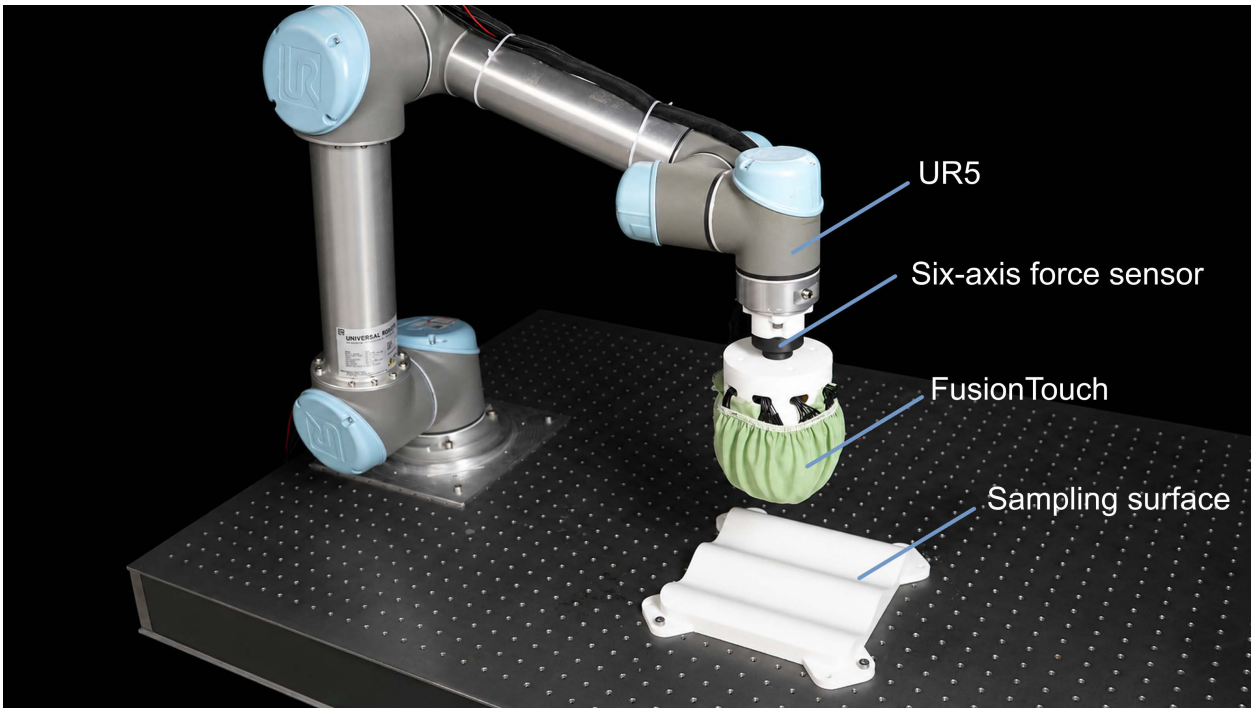


Fig. S23. **Mold-based force-data sampling setup.** The figure shows the experimental setup used to collect force-labeled FusionTouch samples on predefined sampling surfaces. A FusionTouch sensor was mounted at the end effector of a UR5 robotic arm, with a six-axis force sensor installed between the robot wrist and the FusionTouch module. During sampling, the robot drove FusionTouch to contact the molded sampling surface at controlled positions and orientations, while the six-axis force sensor recorded the corresponding contact force. This setup enabled repeatable acquisition of tactile responses and reference force measurements under controlled surface geometries.

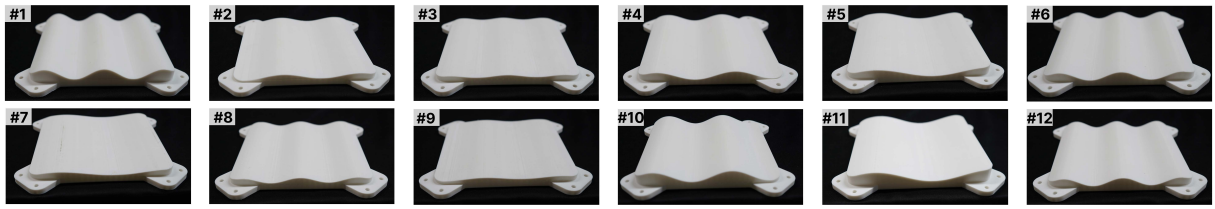


Fig. S24. **Parameterized sinusoidal panels for contact-force calibration.** The twelve panels (#1–#12) were fabricated with different sinusoidal surface profiles, including variations in wave number, amplitude, and phase. During force-calibration data collection, each panel was pressed against the FusionTouch surface by the robot under multiple pneumatic pressure settings, while the internal RGB image, infrared marker image, pneumatic pressure, robot pose, and six-axis force/torque readings were synchronously recorded. These panels generated diverse, repeatable contact geometries for training and evaluating the pressure-conditioned contact-force estimation model.

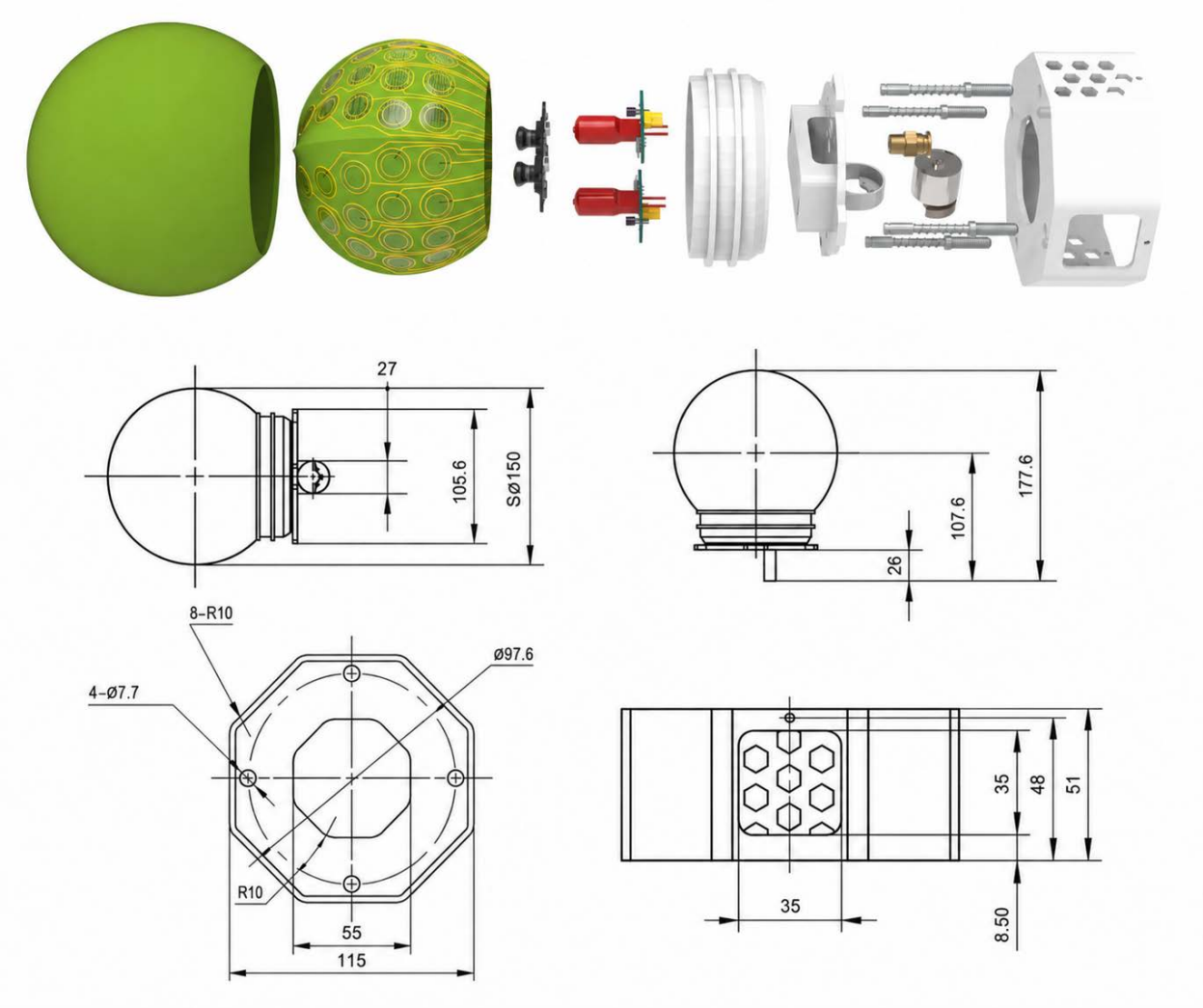


Fig. S25. **Exploded view and key dimensions of FusionTouch.** (A) Exploded view of the FusionTouch assembly, showing the spherical soft shell, patterned electrotactile and optoelectronic skin, internal vision and actuation components, pneumatic and mechanical connection parts, fasteners, and the base housing. (B) Engineering drawings of the assembled device and base structure, showing the major geometric dimensions used for fabrication and assembly. Dimensions are given in millimeters.

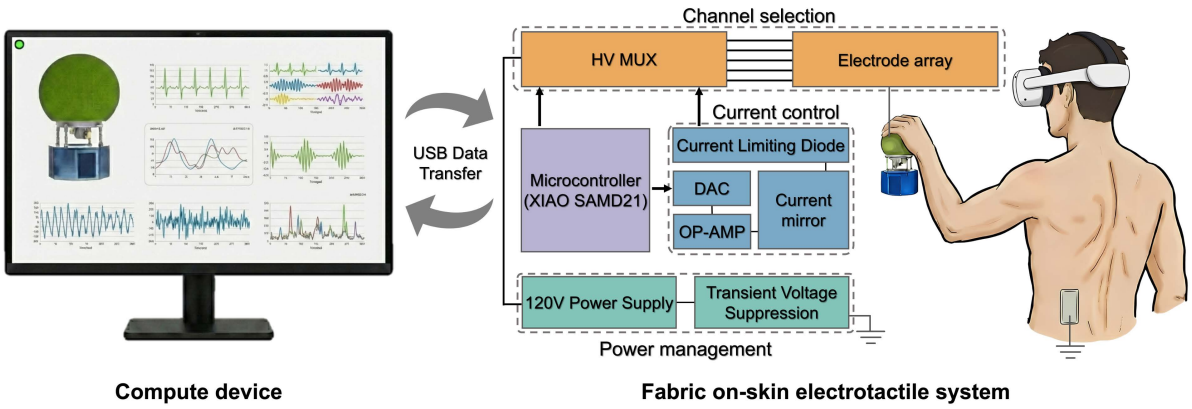
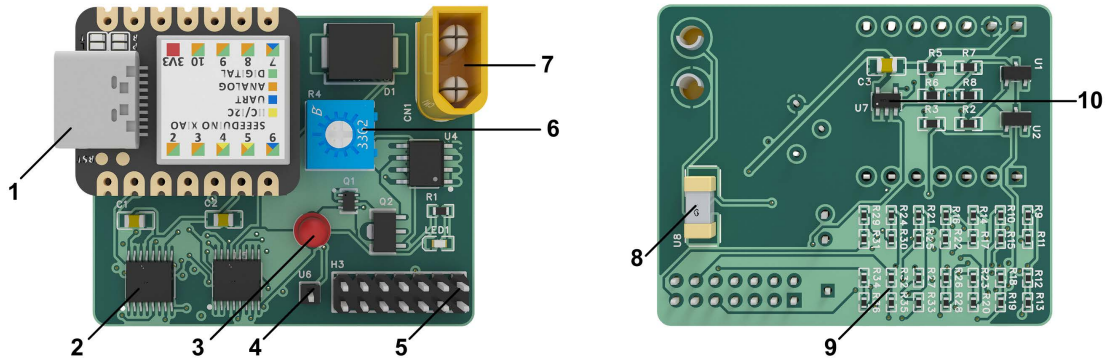


Fig. S26. Hardware architecture and operating principle of the fabric on-skin electrotactile system. The computing device sends stimulation patterns, target channels, and intensity parameters to the embedded controller through USB. The microcontroller (XIAO SAMD21) generates channel-selection and stimulation-control signals, and the high-voltage multiplexer (HV MUX) routes the stimulation output to the selected sites in the fabric electrode array. The current-control stage, consisting of a DAC, operational amplifier, current mirror, and current-limiting diode, regulates and limits the output current to provide programmable and protected stimulation intensity. The power-management stage supplies an approximately 100 V high-voltage source and uses transient-voltage suppression to reduce disturbances from high-voltage pulses. This architecture enables multi-channel, region-selective on-skin electrotactile feedback, allowing virtual or robot-side contact information to be rendered as localized cutaneous stimulation.



1 Main control board **3** Current limiting diode **5** Stimulating electrodes **7** Power supply **9** Safety resistor

2 Multiplexer **4** Public land **6** Adjustable resistor **8** Safety wire **10** OC Comparator

Fig. S27. **Hardware layout of the petal-level electrotactile controller board.** The images show the front and back sides of the PCB used to drive one petal module of the FusionTouch electrotactile array. Each petal is controlled by an independent board that interfaces with the local electrode sites on that petal. The board integrates the main control unit, multiplexers, current-limiting diodes, common ground, stimulating-electrode connectors, adjustable resistor, power-supply input, safety wire, safety resistors, and over-current comparator. After receiving commands from the host computer, the controller selects the target stimulation channel through the multiplexer and regulates the output through current-limiting and over-current-protection circuitry, enabling programmable and region-selective electrotactile feedback within a single petal. Four such controller boards are used to drive the distributed electrode array on the spherical FusionTouch interface.

Supplementary Table

Table S1: Statistics of the maximum spherical-envelope diameter of relaxed right hand across 20 volunteers.

Volunteer ID	Diameter (cm)	Volunteer ID	Diameter (cm)
1	12.01	11	11.33
2	11.38	12	14.75
3	11.42	13	14.06
4	12.13	14	13.74
5	11.43	15	12.01
6	10.80	16	11.81
7	11.46	17	12.14
8	10.84	18	10.79
9	11.41	19	13.64
10	12.55	20	11.63
Mean ± s.d. = 12.07 ± 1.13 cm			

Table S2: Statistics of the maximum spherical-envelope diameter of hand motion during activities. P90 and P95 denote the 90th and 95th percentiles of the frame-wise envelope diameter for each volunteer, respectively. The <15 cm column reports the percentage of frames whose envelope diameter is smaller than 15 cm.

Volunteer ID	P90 (cm)	P95 (cm)	Max (cm)	<15 cm (%)
00000	13.54	13.68	16.53	99.64%
00001	13.34	13.59	15.12	99.75%
00002	12.05	12.38	16.83	99.75%
00003	12.95	13.39	16.25	99.73%
00004	14.33	14.69	17.51	97.34%
00005	13.49	13.93	16.75	99.33%
00006	12.80	13.10	16.87	99.57%
00007	12.91	13.88	20.56	97.80%
00008	13.63	13.69	16.74	99.39%
00009	13.17	13.70	15.55	99.82%
00010	13.16	13.91	17.64	96.98%
00011	13.06	13.59	16.67	99.84%
00012	12.81	13.59	16.45	97.82%
00013	13.08	13.38	15.82	99.58%
00014	14.13	14.27	17.53	99.37%
00015	12.91	13.20	19.59	99.49%
00016	14.27	14.71	16.90	97.38%
00017	13.79	14.53	18.54	97.33%
00018	12.92	13.32	16.97	99.68%
00019	12.99	13.29	17.34	99.47%
Average	**13.27**	**13.69**	**17.11**	**98.95%**

Table S3: Comparison of FusionTouch with representative haptic interfaces. ✓ = supported; × = not supported; – = not reported or uncertain.

System	Structure	Sensing			Feedback				Application		
		Gesture Recognition	Force Sensing	Deformation Sensing	Thermal	Stiffness Control	Electro-tactile	Vibro-tactile	Tele-operation	VR	Massage
AirTouch (*13*)	Tactile Controller	✓	✓	✓	×	–	×	×	✓	×	×
SenseGlove Nova 2 (*14*)	Glove	✓	×	×	×	✓	×	✓	–	✓	×
WeTac (*15*)	Skin Patch	×	×	×	×	×	✓	×	✓	✓	×
OriRing (*16*)	Ring	–	✓	✓	×	✓	×	×	–	✓	×
Haptiknit (*17*)	Sleeve	×	×	×	×	✓	×	×	–	×	×
ATH-Rings (*18*)	Ring	✓	–	✓	✓	–	×	✓	–	✓	×
DOGlove (*19*)	Glove	✓	✓	×	×	–	×	✓	✓	×	×
CL-HMI (*20*)	Skin Patch	✓	✓	✓	×	–	×	✓	✓	✓	×
Haptic Patch (*21*)	Skin Patch	×	✓	–	×	×	×	✓	–	✓	×
Haptic-feedback Smart Glove (*22*)	Glove	✓	✓	✓	×	×	×	✓	–	✓	×
SPETH Glove (*23*)	Glove	–	–	×	×	×	✓	×	–	✓	×
Drone HMI (*24*)	Skin Patch	✓	×	×	×	×	–	✓	✓	✓	×
EStatiG (*25*)	Glove	–	–	×	×	✓	×	✓	–	✓	×
Stretchable Pose Glove (*26*)	Glove	✓	×	✓	×	×	×	×	✓	–	×
Neuromorphic e-skin (*27*)	E-skin Patch	×	✓	–	✓	×	–	×	×	×	×
VET (*28*)	Tactile Interface	–	–	✓	×	–	✓	×	✓	✓	×
Smart Textile Glove (*29*)	Glove	✓	✓	✓	×	×	×	×	–	–	×
MFE (*30*)	Exoskeleton	✓	✓	–	✓	✓	×	✓	✓	×	×
HaMeR (*31*)	Vision-based	✓	×	×	×	×	×	×	–	–	×
FIBHT (*32*)	Haptic Textile	–	×	×	×	×	✓	×	–	✓	×
PhyTac (*2*)	Force Sensor	–	✓	✓	×	×	×	×	✓	✓	×
Ours	**Sphere**	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓

Table S4: Perception Task Results under Different Feedback Conditions

Perception Task	No Feedback	Thermal Feedback On	Electrotactile On	Pneumatic Change	Vibration On
Force Estimation Marker Detection 3D Reconstruction	96.77	96.87	97.22	97.34	96.00
Gesture Estimation	99.91	99.68	99.63	99.86	99.97
Resistance (kΩ)	1.302	1.387	1.302	1.302	1.302

Table S5: Performance of GNN and CNN models for tactile surface reconstruction.

Model	**Mean CD coarse** $\downarrow$	**CD fine** $\downarrow$	**Hausdorff 95%** $\downarrow$	**F-score @** $2mm$ $\uparrow$	**Contact L2** $\downarrow$
GNN (Ours)	0.0028	0.0026 / 0.0017 / 0.0067	0.0807	0.8754	0.0481
CNN	0.0044	0.0029 / 0.0018 / 0.0084	0.0852	0.8512	0.0549

Table S6: Force estimation performance and ablation study with and without pressure.

Method	Setting	Overall MAE ↓	Overall RMSE ↓	MAE_{Fx} ↓	MAE_{Fy} ↓	MAE_{Fz} ↓
CNN encoder	With Pressure	0.1688	0.2960	0.0717	0.1118	0.3228
CNN encoder	No Pressure	0.2182	0.4201	0.0732	0.1081	0.4733
GNN+patch encoder	No Pressure	0.1613	0.3029	0.0707	0.0854	0.3278
GNN+patch encoder	**With Pressure**	**0.0929**	**0.1722**	**0.0432**	**0.0428**	**0.1927**

Bold values indicate the best results for each metric; the bold method denotes our method.

Table S7: Hand-pose estimation performance under different sensing modalities.

Modality	**Overall MAE ↓**	**Overall RMSE ↓**	**Acc. @1° ↑**	**Acc. @5° ↑**
Single modality (RGB)	0.5654	0.9602	84.2167	99.4225
Single modality (Infrared)	0.4484	0.7658	88.5842	99.8355
Dual modality	**0.3569**	**0.5963**	**92.4162**	**99.9464**

Table S8: Per-participant mean completion time in the constrained teleoperation tasks.

Participant	Condition	Hang cup	Scoop rice	Fold towel	Open drawer	Mean
P1	Without feedback	19 s	32 s	23 s	11 s	21.25 s
P1	With feedback	11 s	27 s	10 s	6 s	13.50 s
P2	Without feedback	20 s	34 s	25 s	12 s	22.75 s
P2	With feedback	12 s	28 s	11 s	7 s	14.50 s
P3	Without feedback	21 s	33 s	24 s	13 s	22.75 s
P3	With feedback	13 s	29 s	12 s	8 s	15.50 s
P4	Without feedback	20 s	33 s	24 s	12 s	22.25 s
P4	With feedback	12 s	28 s	11 s	7 s	14.50 s
Average	**Without feedback**	**20 s**	**33 s**	**24 s**	**12 s**	**22.25 s**
Average	**With feedback**	**12 s**	**28 s**	**11 s**	**7 s**	**14.50 s**
Reduction	**Without – With**	**8 s**	**5 s**	**13 s**	**5 s**	**7.75 s**

Table S9: Per-participant first-attempt success counts in the grasp-and-place teleoperation evaluation. Each entry reports the number of successful first attempts over five recorded trials.

Participant ID	Condition	Daily-1	Daily-2	Food-1	Food-2	Food-3	Total
P1	Without feedback	4/5	4/5	4/5	3/5	4/5	19/25
P1	With feedback	5/5	5/5	4/5	4/5	5/5	23/25
P2	Without feedback	4/5	3/5	4/5	3/5	4/5	18/25
P2	With feedback	4/5	4/5	5/5	4/5	4/5	21/25
P3	Without feedback	4/5	4/5	4/5	3/5	4/5	19/25
P3	With feedback	5/5	5/5	4/5	4/5	4/5	22/25
P4	Without feedback	5/5	4/5	4/5	3/5	4/5	20/25
P4	With feedback	5/5	4/5	4/5	4/5	5/5	22/25
Total	**Without feedback**	**17/20**	**15/20**	**16/20**	**12/20**	**16/20**	**76/100**
Total	**With feedback**	**19/20**	**18/20**	**17/20**	**16/20**	**18/20**	**88/100**

Table S10: Daily manipulation tasks included in the volunteer study.

Task ID	Task	Motion type
T01	Screw operation	Rotational manipulation
T02	Object sorting	Pick-and-place
T03	Block stacking	Stacking and placement
T04	Paper cutting	Cutting / shearing
T05	Stapling papers	Pressing
T06	Pouring beans	Pouring / tilting
T07	Folding	Deformable-object folding
T08	Plug insertion	Insertion and alignment
T09	Keyboard typing	Repetitive tapping
T10	Scooping rice	Scooping and transfer
T11	Wrench insertion	Tool insertion
T12	Writing	Trajectory writing
T13	Erasing	Rubbing / erasing
T14	Cable tightening	Rotational tightening
T15	Battery replacement	Assembly / replacement
T16	Screwing	Screw fastening
T17	Chopstick use	Fine pinch manipulation
T18	Placing	Object placement
T19	USB insertion	Insertion and alignment
T20	Tangling Wire	Coiling and fastening

The motion type describes the dominant hand–object interaction pattern involved in each task.

Movie S1. Overview of FusionTouch. This movie introduces the background and motivation of the study, the biomechanically inspired design rationale of FusionTouch, and the overall system architecture. It further summarizes the major functions of the platform, including glove-free hand interaction, multimodal sensing, and co-located haptic feedback, and provides an overview of the representative applications demonstrated in the study.

Movie S2. Three-dimensional reconstruction based on infrared markers. This movie demonstrates the reconstruction of contact-induced surface deformation using near-infrared marker observations on the spherical interface.

Movie S3. Teleoperated manipulation in folding and constrained placement tasks. This movie demonstrates FusionTouch-based robotic teleoperation in two representative manipulation tasks: towel folding and cup hanging. These tasks illustrate the capability of the system to support dexterous object manipulation, spatially constrained placement, and tactile-feedback-assisted control during physical interactions.

Movie S4. Teleoperated manipulation of handheld tools and articulated objects. This movie demonstrates FusionTouch-based robotic teleoperation in electric-drill moving and drawer-opening tasks. These demonstrations illustrate tactile-feedback-assisted manipulation of handheld tools and articulated objects, showing the potential of the system for complex object interactions in real-world environments.

Movie S5. Teleoperated grasping of fragile and daily objects. This movie demonstrates FusionTouch-based robotic teleoperation for grasping a soft peach and a toy. The demonstrations highlight the system's ability to assist manipulation of both fragile objects and daily objects through tactile feedback, enabling more controlled and adaptive grasping behaviors.

Movie S6. Virtual-reality object pickup and manipulation tasks. This movie demonstrates FusionTouch-enabled object manipulation in virtual-reality environments, including apple picking, cup pickup, and steamed-bun pickup tasks. These examples illustrate the application of FusionTouch

in immersive interaction, tactile-feedback-enabled object manipulation, and the handling of both rigid and soft virtual objects.

Movie S7. Virtual-reality dynamic interaction tasks. This movie demonstrates dynamic interaction tasks in virtual reality using FusionTouch, including basketball playing, volleyball playing, and animal interaction. These demonstrations highlight the capability of the system to support immersive, tactile-feedback-enabled interactions with dynamic virtual objects and agents.

Movie S8. Massage experiments in simulated and real human-care scenarios. This movie demonstrates massage experiments using FusionTouch on both a mannequin and a real person. These demonstrations illustrate the system's capability for tactile-feedback-assisted contact interaction in simulated and human-centered physical interaction scenarios, showing its potential for healthcare, rehabilitation, and human-care applications.

Supplementary Data S1. Source data The supplementary data file contains the numerical data used to plot the six characterization curves shown in Fig. 2, panels E–J. The data include force–displacement curves under different pressures, voltage responses under different load resistances and concentrations, current responses over time, optical transmittance at different concentrations, and temperature changes under different applied voltages. Column headers indicate the corresponding variables, experimental conditions, and units.

References and Notes

1. R. Likert, A technique for the measurement of attitudes. *Archives of Psychology* (1932).

2. Y. Tang, *et al.*, Digital channel–enabled distributed force decoding via small datasets for hand-centric interactions. *Science Advances* **11** (4), eadt2641 (2025).

3. R. Archana, P. E. Jeevaraj, Deep learning models for digital image processing: a review. *Artificial Intelligence Review* **57** (1), 11 (2024).

4. O. Ronneberger, P. Fischer, T. Brox, U-net: Convolutional networks for biomedical image segmentation, in *International Conference on Medical Image Computing and Computer-Assisted Intervention* (Springer) (2015), pp. 234–241.

5. N. Wang, *et al.*, Pixel2Mesh: Generating 3D Mesh Models from Single RGB Images, in *ECCV* (2018).

6. B. Ward-Cherrier, *et al.*, The tactip family: Soft optical tactile sensors with 3d-printed biomimetic morphologies. *Soft Robotics* **5** (2), 216–227 (2018).

7. M. Li, L. Zhang, T. Li, Y. Jiang, Continuous marker patterns for representing contact information in vision-based tactile sensor: Principle, algorithm, and verification. *IEEE Transactions on Instrumentation and Measurement* **71**, 1–12 (2022).

8. Y. Wang, *et al.*, Dynamic Graph CNN for Learning on Point Clouds. *ACM Transactions on Graphics* **38** (5) (2019).

9. A. Alspach, K. Hashimoto, N. Kuppuswamy, R. Tedrake, Soft-bubble: A highly compliant dense geometry tactile sensor for robot manipulation, in *2019 IEEE International Conference on Soft Robotics (RoboSoft)* (IEEE) (2019), pp. 597–604.

10. J.-C. Peng, S. Yao, K. Hauser, 3d force and contact estimation for a soft-bubble visuotactile sensor using fem, in *2024 IEEE International Conference on Robotics and Automation (ICRA)* (IEEE) (2024), pp. 5666–5672.

11. W. Fan, M. Yang, Y. Xing, N. F. Lepora, D. Zhang, Tac-vgnn: A voronoi graph neural network for pose-based tactile servoing. *arXiv preprint arXiv:2303.02708* (2023).

12. E. Perez, F. Strub, H. De Vries, V. Dumoulin, A. Courville, Film: Visual reasoning with a general conditioning layer, in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 32 (2018).

13. S. Li, *et al.*, AirTouch: A low-cost versatile visuotactile feedback system for enhanced robotic teleoperation, in *2025 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)* (IEEE) (2025), pp. 15525–15532.

14. SenseGlove, SenseGlove Nova 2, https://www.senseglove.com/product/nova-2/ (2025), accessed: 3 May 2026.

15. K. Yao, *et al.*, Encoding of tactile information in hand via skin-integrated wireless haptic interface. *Nature Machine Intelligence* **4** (10), 893–903 (2022).

16. S. Kang, *et al.*, An 18-g haptic feedback ring with a three-axis force-sensing skin. *Nature Electronics* pp. 1–13 (2025).

17. C. du Pasquier, *et al.*, Haptiknit: Distributed stiffness knitting for wearable haptics. *Science Robotics* **9** (97), eado3887 (2024).

18. Z. Sun, M. Zhu, X. Shan, C. Lee, Augmented tactile-perception and haptic-feedback rings as human-machine interfaces aiming for immersive interactions. *Nature Communications* **13** (1), 5224 (2022).

19. H. Zhang, S. Hu, Z. Yuan, H. Xu, DOGlove: Dexterous manipulation with a low-cost open-source haptic force feedback glove, in *Proceedings of Robotics: Science and Systems* (LosAngeles, CA, USA) (2025), doi:10.15607/RSS.2025.XXI.104.

20. Y. Liu, *et al.*, Electronic skin as wireless human-machine interfaces for robotic VR. *Science Advances* **8** (2), eabl6700 (2022).

21. J.-H. Youn, *et al.*, Skin-attached haptic patch for versatile and augmented tactile interaction. *Science Advances* **11** (12), eadt4839 (2025).

22. M. Zhu, *et al.*, Haptic-feedback smart glove as a creative human-machine interface (HMI) for virtual/augmented reality applications. *Science Advances* **6** (19), eaaz8693 (2020).

23. G. Xu, *et al.*, Self-powered electrotactile textile haptic glove for enhanced human-machine interface. *Science Advances* **11** (12), eadt0318 (2025).

24. C. Yiu, *et al.*, Skin-interfaced multimodal sensing and tactile feedback system as enhanced human-machine interface for closed-loop drone control. *Science Advances* **11** (13), eadt6041 (2025).

25. N. Vanichvoranun, H. Lee, S. Kim, S. H. Yoon, Estatig: Wearable haptic feedback with multi-phalanx electrostatic brake for enhanced object perception in vr. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* **8** (3), 1–29 (2024).

26. M. Park, *et al.*, Stretchable glove for accurate and robust hand pose reconstruction based on comprehensive motion data. *Nature Communications* **15** (1), 5821 (2024).

27. W. Wang, *et al.*, Neuromorphic sensorimotor loop embodied by monolithically integrated, low-voltage, soft e-skin. *Science* **380** (6646), 735–742 (2023).

28. C. Zhang, *et al.*, VET: A visual-electronic tactile system for immersive human-machine interaction, in *2025 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)* (IEEE) (2025), pp. 21590–21595.

29. A. Tashakori, *et al.*, Capturing complex hand movements and object interactions using machine learning-powered stretchable smart textile gloves. *Nature Machine Intelligence* **6** (1), 106–118 (2024).

30. Z. Tang, Y. Guo, C. Xiao, MFE: A multimodal hand exoskeleton with interactive force, pressure and thermo-haptic feedback. *IEEE Robotics and Automation Letters* (2026).

31. G. Pavlakos, *et al.*, Reconstructing hands in 3d with transformers, in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (2024), pp. 9826–9836.

32. K. Yao, *et al.*, A fully integrated breathable haptic textile. *Science Advances* **10** (42), eadq9575 (2024).