

Tactile-integrated FlexiRay: Breaking Planar Limits by Harnessing Large Deformations for Flexible, Full-Coverage, Human-like Multimodal Sensing

Author Information

Affiliations

Yanzhe Wang[†], Haotian Guo[†], Hao Wu[†], Huixu Dong*

Grasp Lab, School of Mechanical Engineering, Zhejiang University, Hangzhou, China

Contributions

Y.W., H.G., and H.W. conceived the idea. H.W., H.G., and Y.W. designed and fabricated the proposed FlexiRay hardware. Y.W., formulated the optimization algorithm for the camera and multi-mirror configuration. Y.W., H.W., and H.G. designed and conducted the experiments. Y.W., H.G., and H.W. collected and analysed the data. Y.W., H.G., and H.W. drafted and proofread the manuscript. Y.W., H.G., and H.W. contributed equally to this work[†]. The study was supervised by H.D..

Corresponding author

*Correspondence to: Huixu Dong (e-mail: huixudong@zju.edu.cn)

Abstract

Integrating tactile sensing into soft grippers holds great promise for safer robotic grasping and enhanced human-robot interactions. However, achieving multimodal, high-resolution sensing remains a significant challenge. While existing visual-tactile sensors offer unparalleled spatial resolution at an affordable cost, they lack sufficient structural compliance and rely on rigid frameworks to stabilize optical paths, hindering non-planar contact perception. To address all relevant research gaps, we introduce FlexiRay, a novel soft gripper combining visual-tactile sensing with the bio-inspired FinRay Effect, characterized by low cost,

high compliance, and dynamic sensory coverage. Combining a flexible substrate, adaptive illumination, and temperature-sensitive materials, FlexiRay replicates five out of the seven core human tactile modalities. A novel multi-mirror optical system, optimized for high coverage despite arbitrary dynamic deformations, enables consistent perception with just a single camera. Furthermore, leveraging a human-inspired deep learning framework, we decouple force, contact location, proprioception, texture, and temperature, achieving a force-sensing accuracy of 0.17 N, a contact localization accuracy of 0.96 mm, and a proprioception accuracy of 0.24 mm, while retaining over 90% effective coverage during dynamic interactions. Beyond enabling stable compliant grasping, FlexiRay's structural adaptability and multimodal perception underpin advanced recognition of non-planar interactions. Coupled with a compliant control strategy and an intelligent teleoperation framework, FlexiRay demonstrates safe and adaptive human-robot collaboration, highlighting its potential to advance service robotics toward greater intelligence, safety, and real-world utility.

Introduction

The human tactile system possesses exceptionally complex perceptual mechanisms, consisting of three principal sensory systems: the cutaneous, kinesthetic, and haptic systems¹. These systems enable the human hand to perceive seven key modalities, namely pressure, contact localization, texture, temperature, vibration, proprioception, and pain². Together, they allow humans to perform various intricate and precise tasks³. Translating this exceptional sensory ability to robots brings about significant benefits, yet significant challenges^{1,4}. Over the past three decades, researchers have explored nearly all forms of sensing, such as resistive^{5,6}, magnetic⁷, pressure-sensitive^{8,9}, capacitive⁶, waveguide-based^{10,11}, acoustic¹², and thermal sensing¹³ et al. A special focus has been placed on the development of tactile sensing in soft robotic grippers^{10,12}, aiming to enable more dexterous and safer environmental interactions via the combination of structural adaptivity and tactile perception. Despite substantial advancements, achieving large-area, high-resolution, and multimodal tactile sensing remains a tremendous challenge. This can be primarily ascribed to two factors: the high production cost of taxel-based measurement methods⁵ and the limitations in spatial

resolution of data-driven computational sensors^{11–14}. Therefore, existing designs often struggle in the trade-off among resolution, coverage, cost-effectiveness, and multimodal sensing capabilities. A comprehensive, human-like tactile perception solution remains an elusive goal.

Visual-tactile sensors (VTS) have emerged as a promising solution to address these challenges by leveraging metal-oxide-semiconductor (CMOS) optical arrays to convert multimodal tactile information into high-resolution, pixel-level images^{15–19}, thereby enabling insights into interactions such as pressure^{16,20} and texture^{21,22}. Most existing VTS systems require stable optical paths to avoid occlusion and perception disturbances, which results in bulky designs and a heavy reliance on rigid structures^{21,23–30}. This rigidity leads to a fixed sensing coverage, presenting significant challenges for integrating VTS with flexible grippers, as it conflicts with the compliance and flexibility inherent to soft robotic systems. Although some studies have achieved integrated robotic finger designs through optical path optimization^{27,31} or camera arrangement adjustments^{24,25,31}, these grippers lack the structural compliance necessary for safe interaction. In particular, their structural rigidity restricts perception to planar contact areas, yielding only local, discrete information on non-planar/irregular objects—reducing their versatility and adaptability in complex environments.

To enhance perceptual capabilities during interaction, previous research has explored ways to improve structural compliance while preserving perceptual coverage. For instance, Insight³² and GelSight Svelte³¹ have attempted to introduce a hollow skeletal framework and a curved mirror backplane with material-based elasticity, respectively, yet they allow only limited deformation. Multi-segment fingers with dedicated cameras balance dexterity and perception^{4,33}, though each phalanx remains fundamentally rigid. Likewise, systems like JamTac³⁴ achieve high compliance through jamming, but this mechanism inherently restricts their grasp range and load capacity, narrowing their practical utility. In comparison, bio-inspired FinRay Effect (FRE) grippers enable adaptive and stable grasping via passive structural deformation³⁵. The integration of VTS with compliant FRE architectures offers a promising direction to mitigate the stiffness typical of traditional VTS^{36,37}. Recent advances in this area include replacing rigid acrylic with thin Mylar

films^{36,37}, as well as monolithic silicone molding designs³⁸ and spatial polyhedral networks⁴¹, all achieving improved structural compliance. However, these designs still struggle with visual occlusion and loss of perceptual field under large deformations. To preserve optical functionality, many existing designs deliberately stiffen flexible structures, sacrificing the inherent softness of FRE structures. As a result, embedding VTS within highly compliant structures, without compromising high-resolution and large-area tactile perception capabilities under large structural deformation remains an unresolved challenge.

We present FlexiRay, a sensorized soft gripper that replicates five of the seven essential tactile perceptions in human touch. It passively adapts to objects, stably capturing fine contact textures while simultaneously perceiving force, contact location, proprioception, and temperature, even under substantial structural deformation (Fig. 1). This synergistic design, incorporating the flexible FRE structure, VTS, and a multi-mirror array, not only achieves a sensing area of up to 1560 mm² but also ensures stable perceptual coverage (87.2% on average) across varying deformation states. Inspired by the efficient neural processing of the human tactile system, we have constructed decoupled multi-modal networks that enable multi-modal perception with an end-to-end latency of ~25.8 ms. It further delivers resolutions of 0.17 N (force), 0.96 mm (contact localization), and 0.24 mm (proprioception). In terms of mechanical performance, FlexiRay exhibits over 400% higher deformation capacity than comparable flexible VTS designs, along with excellent durability exceeding 40,000 load cycles. A detailed comparison between the representative state-of-the-art VTS sensors and FlexiRay is summarized in Table 1.

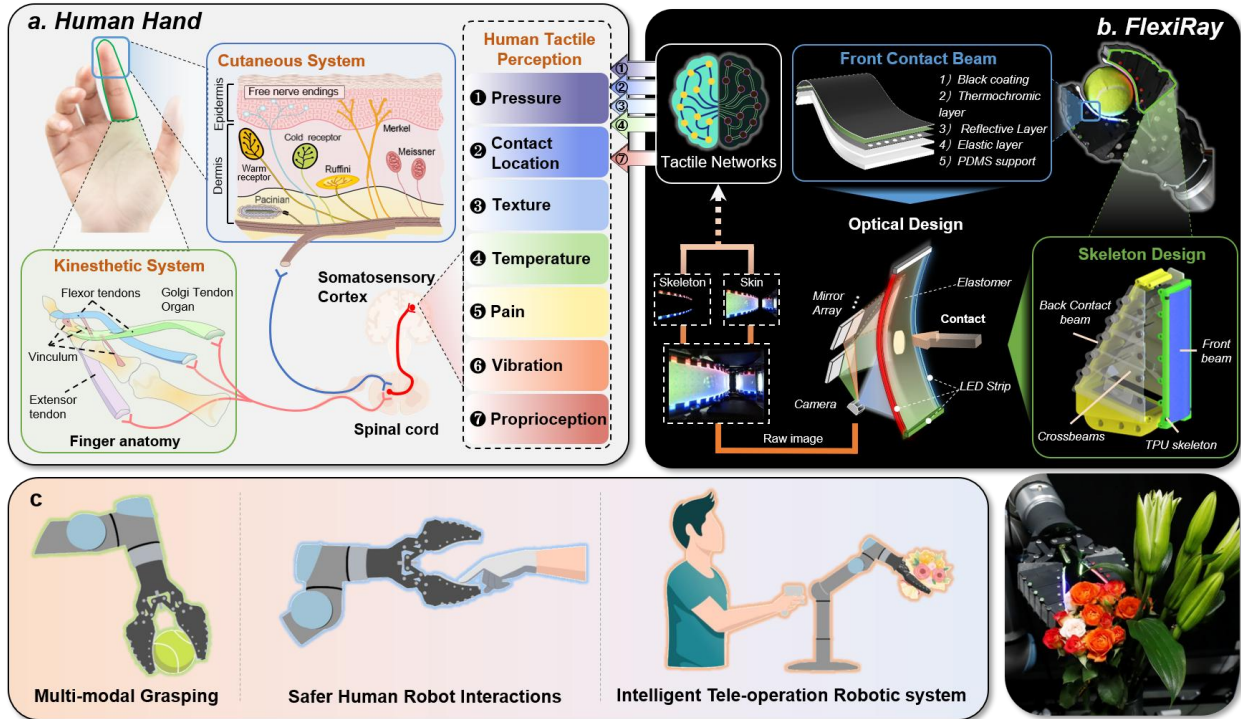
Main contributions are summarized as follows:

- We present a system-level framework merging visual-tactile sensing with compliant mechanics, exemplified by FlexiRay, a soft gripper realizing five tactile modalities (force, contact location, texture, temperature, proprioception). FlexiRay preserves structural compliance for safe interaction while ensuring robust multi-modal tactile sensing under large-scale passive deformation.
- A multilayer structure is developed by integrating an FRE substrate with a PDMS-based contact layer, combined with thermosensitive materials and soft silicone tactile materials. This configuration

replaces the rigid skeleton in conventional VTS, enhancing system durability exceeding 40,000 load cycles while expanding the perceptual range to incorporate temperature sensing.

- An adaptive strategy based on covariance matrix adaptation evolution optimizes the internal optical system. By leveraging deformation-induced structural variations, the method maximizes dynamic sensing coverage using a single camera and multiple mirrors. Optical interference is transformed from a limiting factor into a functional design element, while distributed mirrors ensure continuous perception during deformation without compromising compliance.
- Exploiting FlexiRay's high compliance and wide sensing coverage for real-world tasks, a multi-modal perception tactile network is established. Demonstrations include compliant grasping of fragile and soft objects (e.g., chips, cake), sensitive and safe human–robot interaction (e.g., feather-triggered responses, tactile-guided kinesthetic teaching), adaptive grasping, classification of diverse objects, and multi-modal feedback enhanced real-time multimodal feedback (force, position, proprioception, temperature) for teleoperation. These capabilities expand the possibilities of dexterous manipulation and intelligent collaboration in unstructured environments.

Fig. 1: FlexiRay: A flexible FinRay Effect gripper enables human-like multi-modal tactile and proprioceptive perception.



a The human haptic system integrates kinesthetic and cutaneous perceptions to enable dexterous action. Kinesthesia, mediated by muscles, tendons, and joints, conveys real-time posture and motion for fine tasks. Cutaneous receptors, including Pacinian corpuscles, Merkel's discs, Ruffini endings, Meissner's corpuscles and temperature receptors, sense object properties and interaction dynamics. Together, these multi-modal perceptions interpret contact characteristics to support dexterous manipulation. **b** Embodying the FinRay Effect, FlexiRay offers exceptional structural compliance. Its novel design incorporates a multi-functional sensing pad on the front contact beam, alongside an optimized internal optical system where a single camera and multi-mirror array convert optical interference into usable perceptual signals. This integration ensures continuous sensing throughout deformation. Additionally, a learning-based multimodal perception model replicates five of the seven perceptual modalities found in the human hand. **c** The human-like multimodal perception and structural compliance of FlexiRay enable safe and effective application in multimodal compliant grasping, human-robot collaboration, and intelligent teleoperation.

127 Table 1 Comparison between the state-of-the-art VTS systems and FlexiRay.

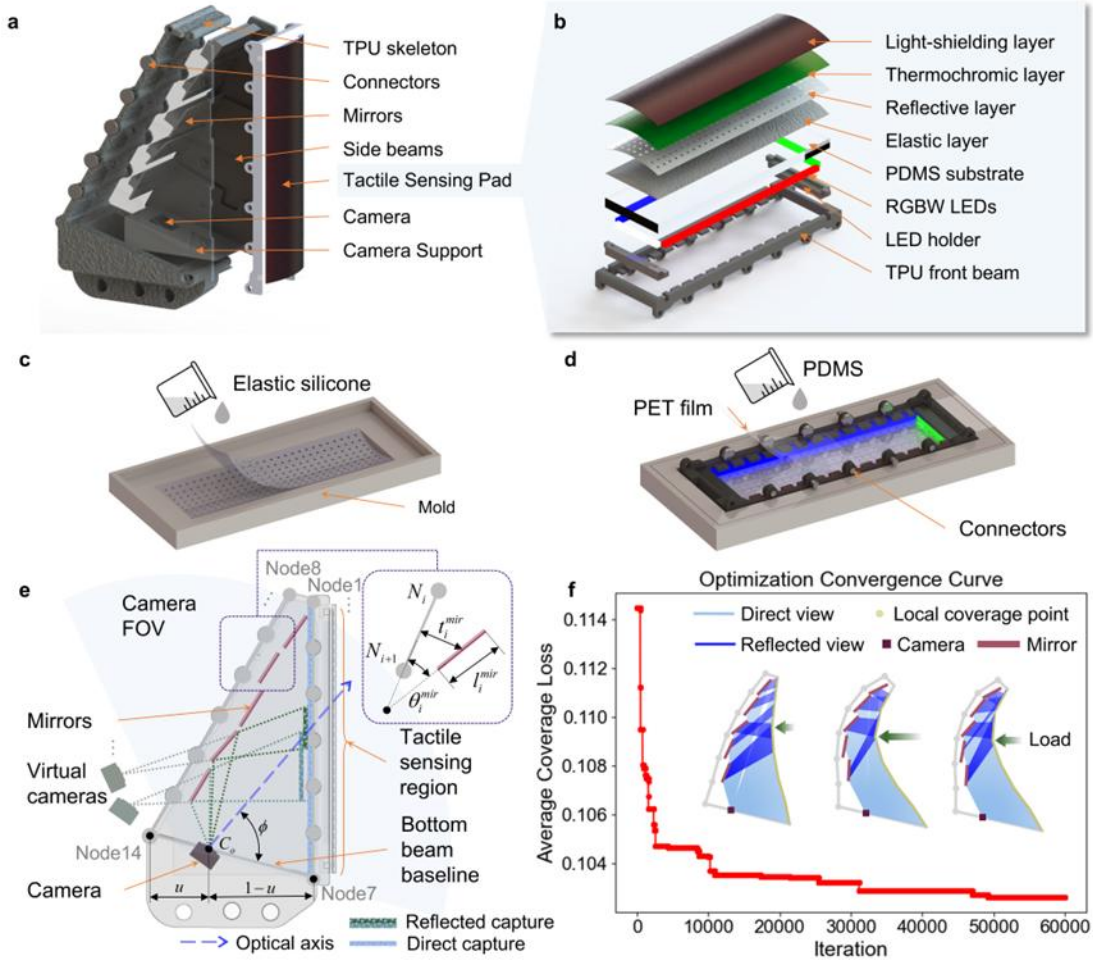
Sensor	Principle	System Rigidity	Area (mm ²)	Method	Modality	Force Error (N)	Location Error (mm)	Proprioception Error	Non-planar Adaptability
GelSight ²¹	Camera	Rigid	250	CNN	Force, Texture	$F_N: \sim 0.67$	—	—	Limited contact area
GelSight360 ²⁵	Camera		—	MLP	Texture	—	—	—	
Digit ²⁶	Camera		304	ResNet	Texture	—	—	—	
Gelslim ²⁷	Camera		1200	iFEM	Force, Texture	$F_N: \sim 0.32$	—	—	
GelSight Wedge ³⁹	Camera + Single-mirror		768	MLP	Texture	—	—	—	
HaptiTemp ¹⁵	Camera		—	CNN	Texture, Temperature	—	—	—	
DigiTac ²⁸	Camera		1256	PoseNet	Force, Location	—	0.6 - 1.2	—	
GelTip ³⁰	Camera	Rigid finger base	—	Perspective Model	Location	—	5	—	
TacPalm ²⁹	Camera	Rigid palm	565.5	ResNet	Texture	—	2.3	—	
Insight ³²	Camera	Semi-compliant (Hollow skeleton)	4,800	ResNet	Force, Texture	0.03	0.4	5°	
GelSight Svelte ³¹	Camera + Single-mirror	Semi-compliant (Soft front beam)	~1895	CNN	Bend/ Twist, Texture	$T_B: \sim 9.4$ Nmm/ $T_T: \sim 7.6$ Nmm	—	—	For each phalange, limited contact area
GelFlex ⁴	Multi-Cameras	Highly flexible (Serial Joints)	—	LeNet	Proprioception, Texture	—	1.71	~0.77 mm	
GelSight Endoflex ⁴⁰	Camera	Highly flexible (Soft Endoskeleton)	—	ResNet	Texture	—	—	—	
GelLink ³³	Cameras + Multi-mirrors	Highly flexible (Serial Joints)	3060	Geometric Optimization	Texture, Location, Proprioception	—	—	2.3°	
JamTac ³⁴	Camera	Soft (Jamming)	<6358.5	PSA&CNN	Force, Location, Texture	0.3	6.0	6.0 mm	Limited gripping surface

Liu et al. ⁴¹	Camera	Compliant (Soft Polyhedral Network)	—	MLP	Force, Proprioception	F_N : ~0.25	—	~1.18 mm	Only perceive contact at the beams
GelSight Baby Fin Ray ²²	Camera + Single-mirror	Compliant (FRE with rigid connections)	990	ResNet	Texture	—	3.6 pixels	—	Sacrifice structural compliance to maintain stable optical paths
AllTac Fin Ray ³⁸	Camera	Compliant (FRE with flexible connections)	>1120	Diff-Mapping & CNN	Force, Location, Proprioception	0.244	<1	3.22°	Limited coverage when deforming
FlexiRay (Ours)	Camera + Multi-mirrors	Compliant (FRE with flexible connections)	1560	ResNet + PP-LiteSeg	Force, Location, Proprioception, Texture, Temperature	F_N : ~0.17	0.96	~0.24 mm	Average coverage of 87.2% under different deformations

Results

Working principles of FlexiRay

Fig. 2: Illustration of the integrated design and layout optimization of FlexiRay.



a Exploded view of FlexiRay. **b** Exploded view of the tactile sensing pad. **c** Casting of the elastic layer. **d** Casting of the PDMS substrate. **e** Schematic diagram of the inside optical sensing system layout optimization method. **f** The optimization convergence curve of average coverage loss.

The design of the proposed Tactile-integrated FlexiRay is illustrated in Fig. 2a. The system is primarily composed of a compliant finger framework, an optical system, and a tactile sensing pad. The compliant

finger framework includes a back beam, a front beam, and side beams. The back and front beams, made of Thermoplastic Polyurethane (TPU) material, provide elasticity similar to the tendons in the human hand, maintaining structural stability while transmitting forces and deformations. The rigid side beams act as the bones of the hand, being hinged to the ends of the back and front beams to provide movable joints. The optical system includes a camera for image capture, multi-segment reflective mirrors for enhancing the camera's field of view (FOV), and flexible LED light strips for illumination. The tactile sensing pad consists of a PDMS-based substrate, an elastic silicone layer, a reflective layer, a temperature-sensing layer, and a light-shielding layer, as shown in Fig. 2b. The PDMS substrate provides flexible support without compromising compliance (Fig. 2d), while the low-hardness transparent silicone elastic layer is used to enable texture mapping (Fig. 2c). The external silicone reflective layer, blended with aluminum powders, enables fine contact texture recognition. Atop this reflective layer is the temperature-sensitive layer embedded with thermochromic materials, whose thermally induced changes in the markers can be captured by the camera. The outermost light-shielding layer is responsible for mitigating external illumination interference. Thus, FlexiRay exhibits sensitivity to external physical contacts and temperature stimuli, similar to the mechanoreceptors and thermoreceptors in human skin.

The key challenge in integrating VTS with flexible structures lies in the significant deformation of the soft robotic gripper during the interaction, which obstructs the camera's FOV. Rather than limiting the flexibility of the hand, we address this issue through systematic optimization, incorporating a multi-mirror configuration to transform the negative structural deformation into a functional advantage. As shown in Fig. 2e, each mirror is independently attached to the back beam, passively altering the direction of the camera light path. This allows the discrete capture of perception regions that the camera's FOV does not cover. Combining the views from multiple mirrors achieves continuous coverage of a large perception area under dynamic loading conditions. To ensure optimal compliance adaptation, the optical system layout is modeled as a 2D geometric parameter optimization problem in the lateral cross-section. The optimization objective is maximizing the coverage of the tactile sensing regions captured by direct camera views and mirror

reflections. The optimization parameters include the camera’s position along the bottom beam baseline, its shooting angle, the length of each mirror, as well as the distance and angle of each mirror relative to the back beam. CMA-ES is used as the solving tool. Deformation data for the back and front beam joint nodes under different loads are collected to serve as the prior information. This optimization process essentially aims to find the layout parameters that maximize the camera's FOV coverage across various deformations, leveraging the passive FOV reconstruction capability of multiple mirrors to enhance the camera’s perception of blind regions in deformed views. The convergence curve of the coverage loss during the optimization process is shown in Fig. 2f.

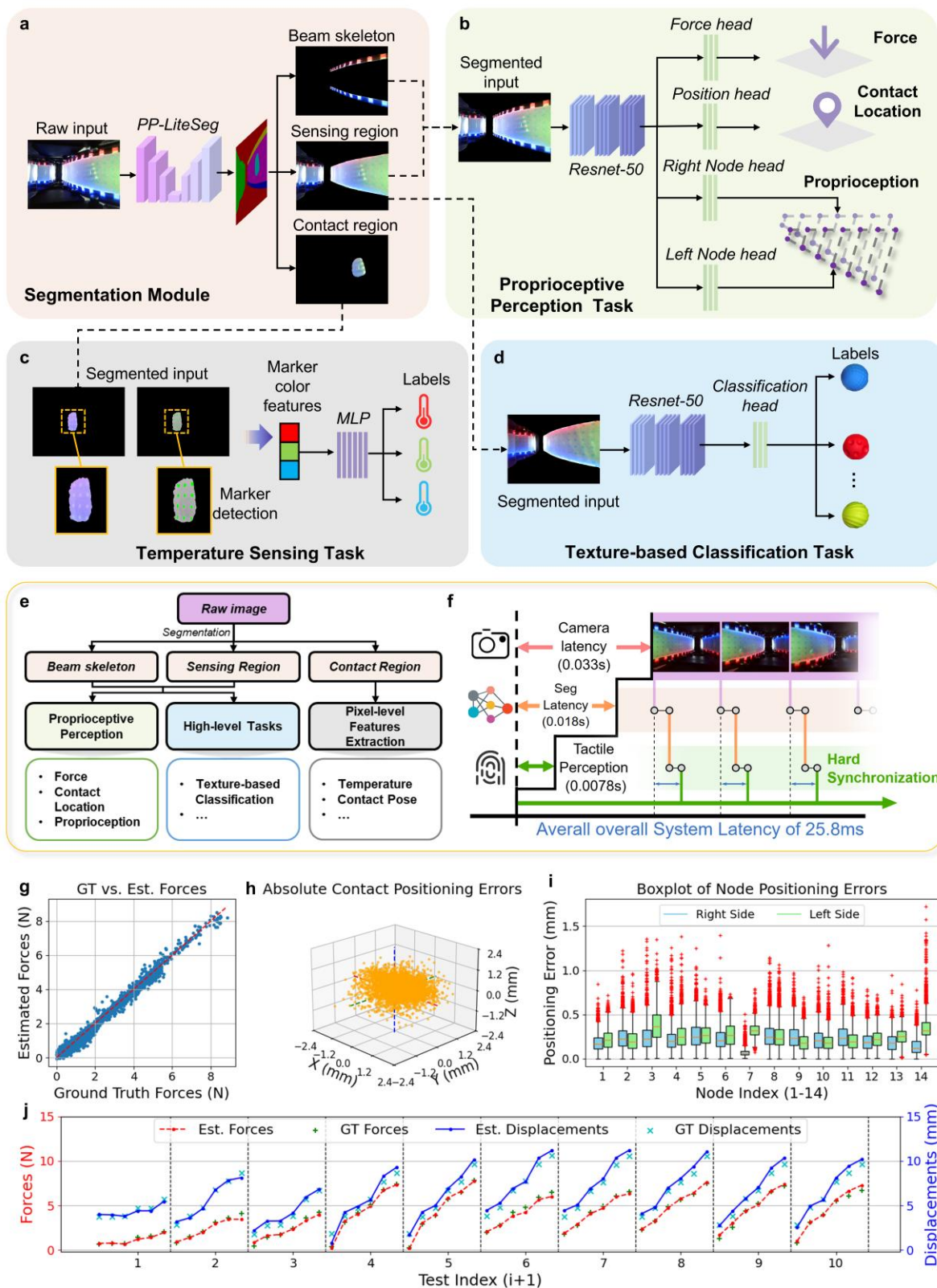
Multi-modal perception model construction and performance evaluation

To achieve multi-modal perception in 3D space, we develop a series of deep learning models that decouple perception tasks and allocate them to specialized models. This approach effectively addresses the challenging spatial deformation coupling between the flexible base and the contact area during interactions. Additionally, combining these sub-models enables the solution of more complex real-world tasks. First, we developed an image region extraction model based on the PP-LiteSeg model (Fig. 3a). This lightweight semantic segmentation model aids in distinguishing contact information into cutaneous and kinesthetic systems, segmenting the front beam skeleton, the front beam interaction region (direct perception), the mirror region (reflective perception) as well as the local contact region, serving as a preprocessing step to enhance the quality of subsequent tasks while eliminating environmental interference. Next, a proprioception model based on the ResNet50 backbone is implemented (Fig. 3b), with the segmented images of the front beam skeleton and the sensing region as input. This model estimates the normal contact force, the 3D position of the contact point relative to its reference frame, and the side beam node positions in the global coordinates, using dedicated heads for force, position, and node estimation. Furthermore, FlexiRay’s multimodal perception enables high-level tasks, such as fine texture recognition. A texture recognition model, built on a ResNet50 backbone (Fig. 3d), takes sensing region images as input and classifies contact textures via a classification head. Additionally, a temperature sensing model (Fig. 3c) is

separately developed, which utilizes the color features in the markers of the local contact region image to provide contact temperature information.

Figure 3g illustrates the estimated normal contact forces across 5,000 distinct contact locations and loading conditions. The horizontal axis represents ground-truth measurements, while the vertical axis shows model predictions. The predicted forces achieve a root-mean-square error (RMSE) of 0.17 N and a correlation coefficient of 0.995, indicating near-perfect agreement with physical measurements and underscoring the high fidelity of the force-sensing model. Figure 3h presents the distribution of absolute errors in estimated 3D contact positions under the same 5,000 loading conditions. The closer a point lies to the origin, the smaller the error. Statistical analysis yields a mean error of 0.96 mm with a standard deviation of 0.45 mm, confirming both the accuracy and stability of the model in spatial contact localization. For proprioceptive representation of finger configuration, Figure 3i shows box plots of localization errors for 14 sensing nodes on each side of FlexiRay under the same loading conditions. The right-side nodes exhibit a mean localization error of 0.21 mm (mean s.d.: 0.13 mm), while the left-side nodes yield 0.26 mm (mean s.d.: 0.15 mm). These results demonstrate consistently high spatial precision and robustness. To further evaluate proprioceptive accuracy during continuous interaction, we conduct tests where contact depth was incrementally increased from a randomized initial position to a target load. Figure 3j compares measured and estimated values of both normal force and contact depth over 10 repeated randomized trials. The model maintains high prediction fidelity and stability even under dynamic, continuous loading, validating its robustness in real-time proprioceptive sensing. The image processing pipeline is shown in Fig. 3e. Camera capture (30 fps, 0.033 s), segmentation (0.018 s), and proprioceptive inference (force, contact localization, proprioception; 0.0078 s) yield an end-to-end latency of ~25.8 ms. Figure 3f illustrates perception latency and alignment, where all modalities share the same image stream to ensure temporal consistency without additional post-processing, enabling low-latency, synchronized multimodal perception for dynamic interactions.

213 **Fig. 3: Learning-based multi-modal perception pipeline and proprioception accuracy analysis.**



a Semantic segmentation model is employed to segment the front beam skeleton, perception region, and contact region. **b** Sub-model for sensing normal contact force, position, and proprioception deformation. **c** Color mapping model for temperature sensing. **d** Texture recognition model for tactile-based object classification. **e** Image-processing pipeline of FlexiRay. **f** Schematic visualization of perception latency and temporal alignment. **g** Accuracy analysis of normal contact force estimation under 5000 varying loading trials. **h** Absolute error distribution of contact position prediction under 5000 varying loading trials. **i** Box plot of positioning errors for joint nodes under 5000 varying loading trials. **j** Continuous estimation of normal contact force and depth under 10 random tests. "Est." refers to the estimated data, and "GT" refers to the ground truth data.

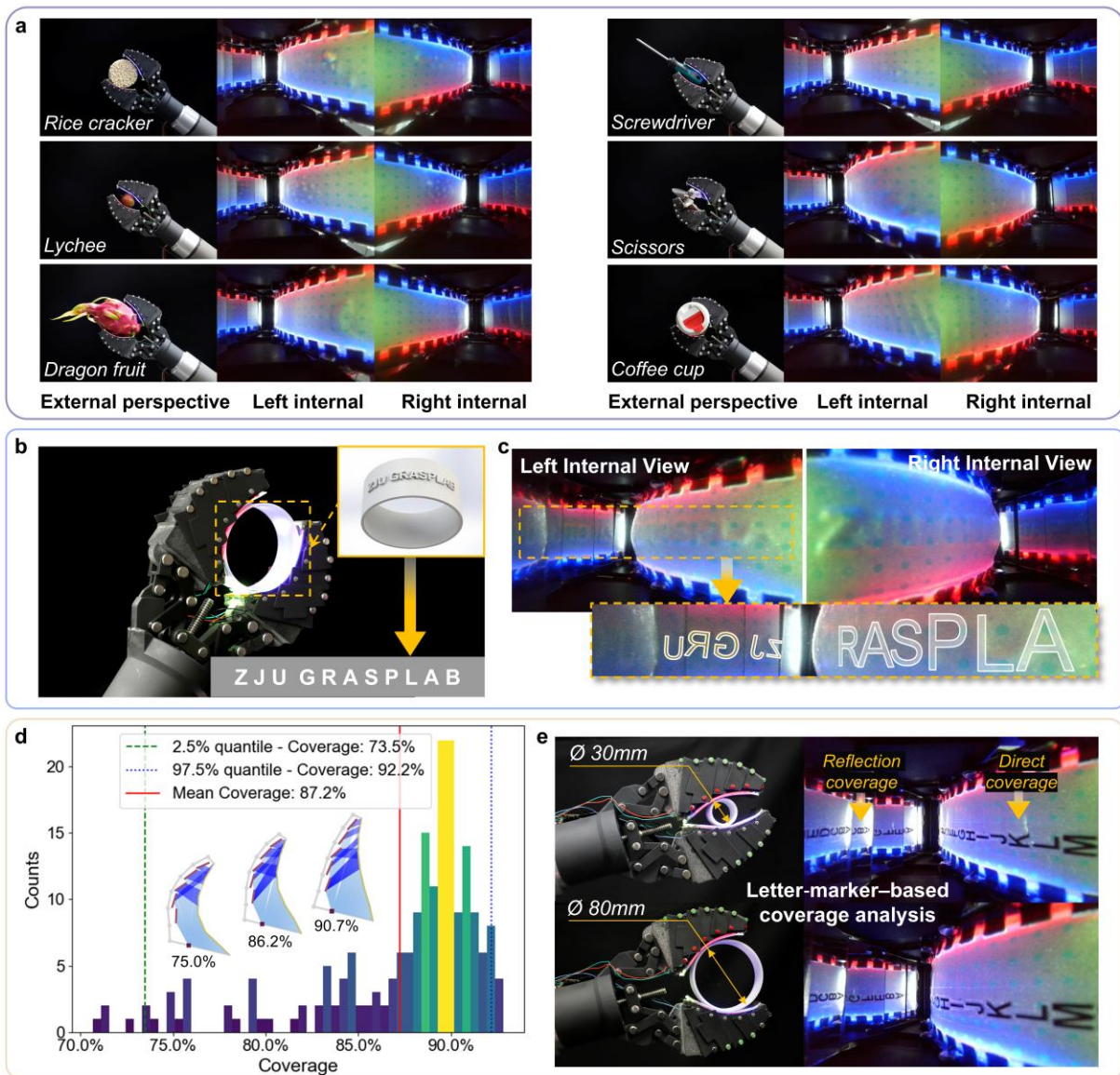
Exquisite texture detection and perceptual coverage

To assess FlexiRay's texture detection performance, a series of test objects are selected, encompassing large curved or large-area items (e.g., screwdrivers, wire strippers, scissors), and curved objects of varying sizes and shapes (e.g., solder, bottle caps, vases, pen holders, clips, mice). A two-fingered gripper composed of FlexiRay is mounted on a UR5e robotic arm to perform natural grasping experiments and capture texture images without external interference. Several typical demonstrations are shown in Fig. 4a, with the complete set provided in the Supplementary Fig. S4. The results demonstrate that FlexiRay not only conforms seamlessly to and wraps around large curved objects but also accurately captures the surface contours and geometric details during flexible deformations. This showcases the extensive tactile perception capability of FlexiRay, which is not available in current VTS technologies.

The practical effect of the multi-mirror configuration on FOV extension is evaluated using the 3D-printed text ring shown in Fig. 4b. The ring has an outer diameter of 68 mm, a font height of 1.5 mm, and a line width of 1 mm. As shown in the internal view in Fig. 4c, under large deformations, the front beam obstructs the camera's capture of the fingertip region. However, by incorporating mirror-reflective areas, FlexiRay is able to capture the occluded textures, thus providing a more complete perceptual field. Specifically, the

direct perception area captures "RASPLA," while the reflective perception area captures "ZJUGR," forming a continuous texture pattern "ZJUGRASPLA."

Fig. 4: Texture detection performance and perceptual coverage of FlexiRay.



a Raw internal images captured while gripping various objects. **b** 3D-printed text ring. **c** Internal raw image of the text ring grip, illustrating texture details captured by both direct and reflective sensing. **d** Distribution of perceptual coverage across 200 random deformation tests under varying loads. **e** Continuous alphabetic markers indicate the visible areas during ring-shaped grasps of different diameters, with the 30 mm and 80

mm rings shown as representative examples. Additional demonstrations are provided in the Supplementary Fig. S6.

The perceptual coverage under different deformations is further quantified through both simulation and physical experiments. The simulation is conducted based on the collected hand configuration data at 200 random contact positions and loads. The camera's FOV is uniformly discretized into 300 rays in a 2D plane, and the coverage of each ray on the contact area is calculated, including both direct incidence and mirror-reflected coverage. The contact area is discretized into 100 target points. Considering the varying propagation distance l^c of each ray, the coverage radius of a single ray is defined as $R = l^c \sqrt{2(1 - \cos \Delta\theta)}$, where $\Delta\theta$ is the angle between two adjacent rays. The perceptual coverage is assessed by calculating the proportion of target points covered by each ray. Experimental results (Fig. 4d) show 87.2% average contact-area coverage across all 200 random tests, with most of the coverage concentrated around 90%. Even under extreme deformations, the perceptual coverage remains above 70.0%. In particular, physical experiments further evaluate practical perceptual coverage stability under varying structural deformation. Rings with diameters ranging from 30 to 80 mm are chosen to mimic most real-world contact scenarios. The continuous alphabetic markers on the perceptual area help visually track coverage variations as deformation changed. Results for the smallest (30 mm) and largest (80 mm) diameters are shown in Fig. 4e, with remaining cases provided in the Supplementary Fig. S6. Throughout all compliant grasping tests, FlexiRay consistently yields complete alphabet sequences, confirming near-full contact-surface coverage even across different deformation degrees.

Texture-based object classification during compliant grasping

FlexiRay's large-area perception and fine texture sensing capabilities also facilitate more efficient and robust object classification during compliant grasping. This was evaluated by using eight textured sensory balls (diameters: 61–82 mm) as classification targets and comparing it with the state-of-the-art commercial VTS, GelSight Mini, which serves as the baseline. For each target, 80 tactile images are randomly acquired

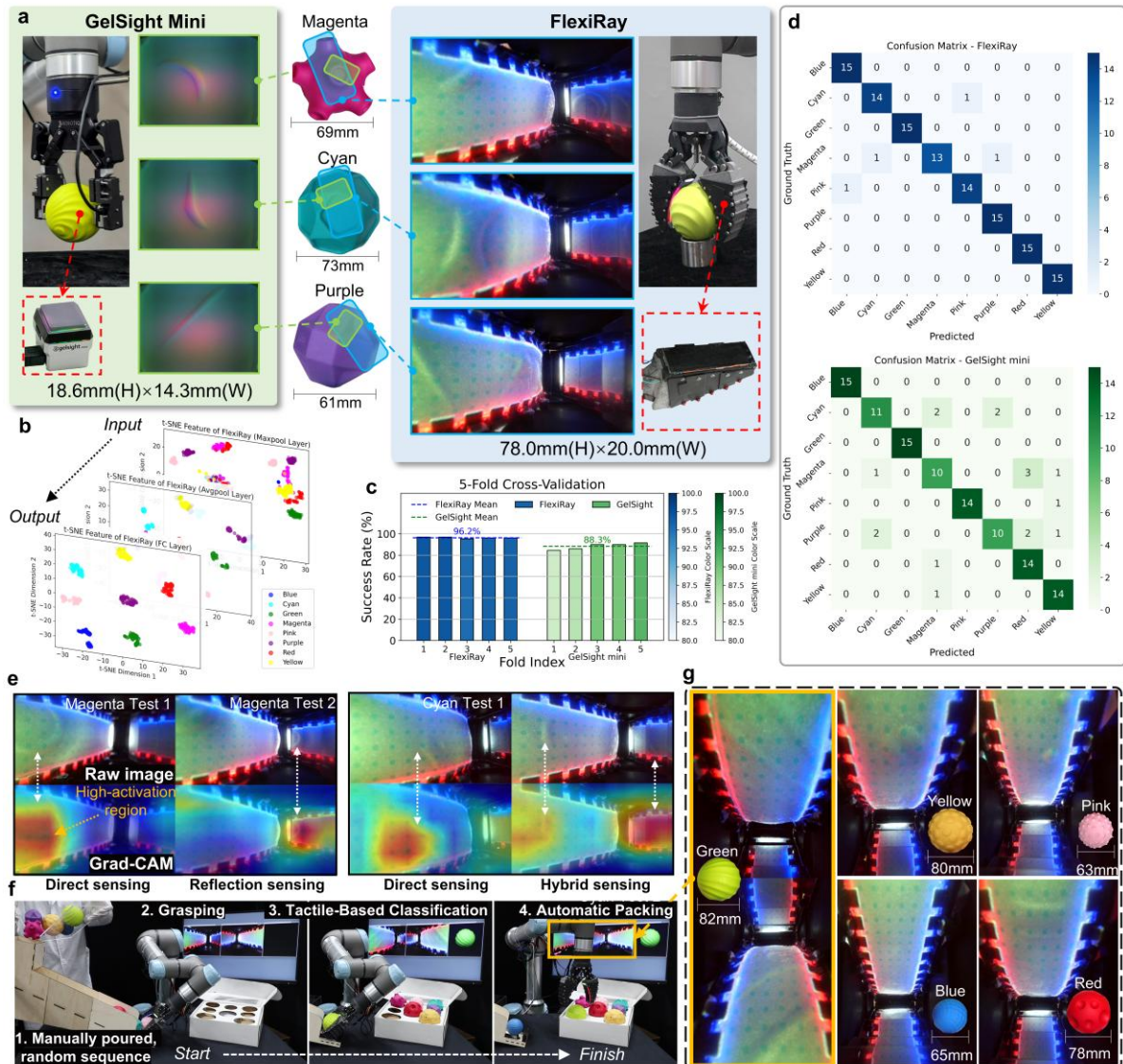
during stable grasps, yielding a total of 640 samples. The dataset is split into training and validation sets at a 4:1 ratio, and five-fold cross-validation is performed to mitigate potential distribution bias. The GelSight Mini sensor is employed under identical sampling, training, and parameter settings. As shown in Fig. 5c, FlexiRay achieves an average validation accuracy of 96.2% across folds, outperforming GelSight Mini’s 88.3% accuracy. The narrow variance across folds for both sensors indicates balanced dataset distribution. To highlight feature separability, Fig. 5b presents t-SNE visualizations of features from the best-performing FlexiRay model, extracted sequentially from the max-pooling, average-pooling, and fully connected layers, demonstrating both effective model design and training efficacy.

For independent evaluation, each gripper grasps every ball 15 times, generating two test sets of 120 samples each. As shown in the confusion matrices (Fig. 5d), FlexiRay achieves an average classification accuracy of 96.7%, with perfect accuracy for most balls and the lowest (86.7%) for the magenta ball, which is occasionally misclassified as cyan or purple due to similar edge textures. In contrast, GelSight Mini achieves an average of 85.8%, with markedly lower accuracy for magenta, purple, and cyan balls (66.7%, 66.7%, and 73.3%, respectively). These results underscore FlexiRay’s advantage: its sensing area (1560 mm²) is 5.9× larger than GelSight Mini’s (265.98 mm²), and its compliance enables single-contact conformation to curved surfaces, capturing non-planar features at greater distances. This capability even allows simultaneous detection of the magenta ball’s dual protrusions—impractical for GelSight Mini—leading to richer tactile cues and superior recognition performance.

Using gradient-weighted class activation mapping (Grad-CAM) visualizations (Fig. 5e), we identify the regions most influential in classification. For instance, Magenta Test 1 leverages direct illumination over tactile contact regions, whereas Magenta Test 2 focuses on specular reflection zones. Similarly, Cyan Test 1 relies primarily on direct-illumination cues, while Cyan Test 2 incorporates both reflection-dominant and direct-contact features. These examples highlight the importance of specular reflection sensing in soft visuotactile perception, enabling activation of distinct decision regions under varying contact configurations.

Building on the trained model, we implement a robotic ball-packing task (Fig. 5f). Balls are randomly mixed and dispensed onto an inclined feed channel, from which the robot grasps them sequentially. Using only tactile modality, the robot identifies surface textures and sorts balls into designated packing zones. Representative tactile images captured during sorting are shown in Fig. 5g. The task demonstrates FlexiRay's compliance, grasp stability, and robustness in handling varied complex surface geometries, while relying exclusively on tactile sensing for accurate classification, underscoring its role in precise grasping and manipulation.

Fig. 5: Texture recognition and classification in compliant grasping with FlexiRay.



a The comparison of perception modes and coverage states between GelSight Mini and FlexiRay. **b** t-SNE visualizations of features from the best-performing FlexiRay model, extracted sequentially from the max-pooling, average-pooling, and fully connected layers. **c** Five-fold cross-validation accuracies for FlexiRay and GelSight Mini. **d** Confusion matrices for the test results of GelSight Mini and FlexiRay. **e** Grad-CAM visualizations highlighting decision-critical regions under varying contact configurations, revealing the role of specular reflection cues in soft visuotactile perception. **f** Robotic ball-packing task using tactile-only classification to sort balls into designated bins. **g** Raw tactile images of various ball types.

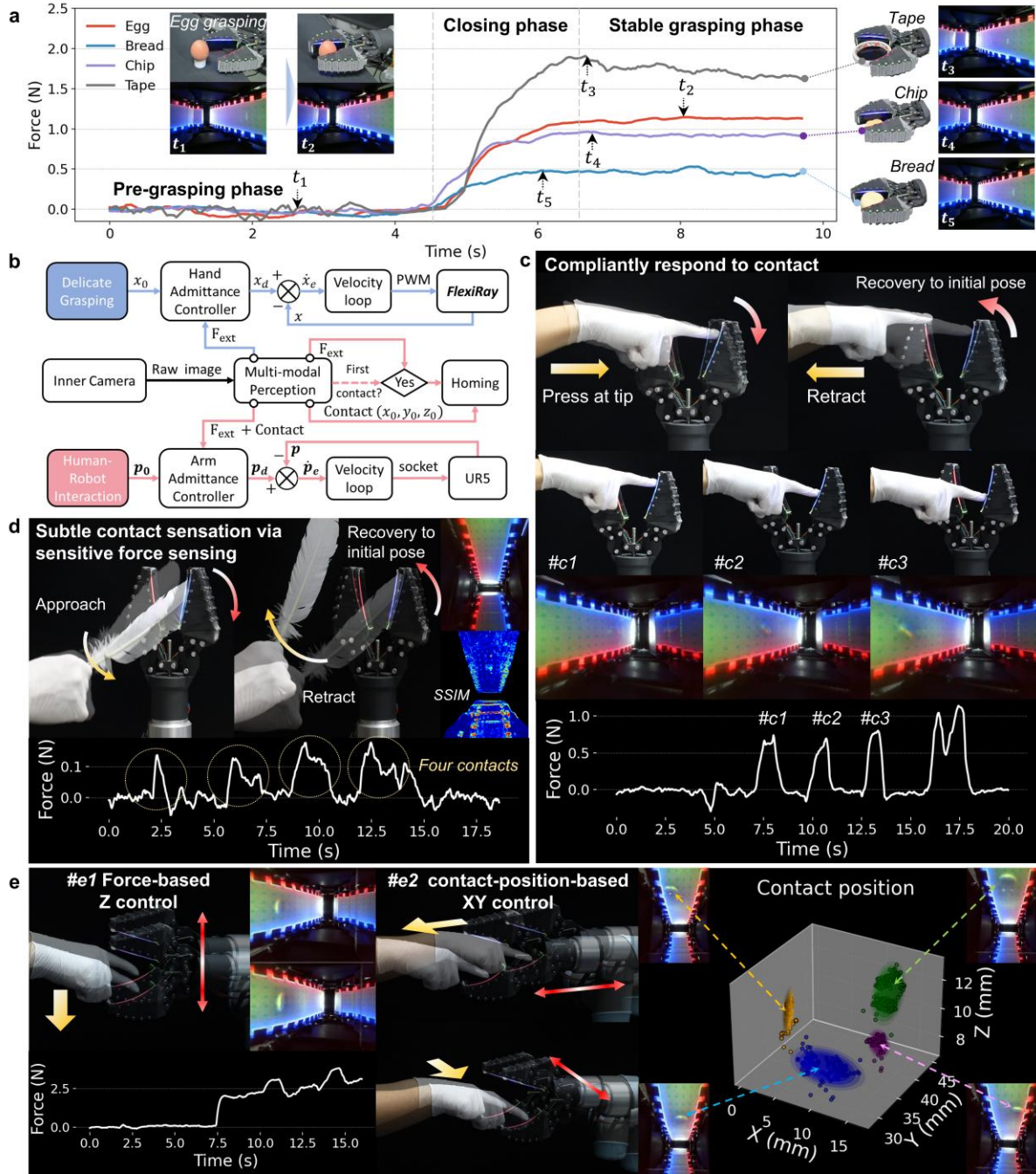
Compliant grasping, safe contact response and force-position sensing for interactive robot control

Grasping deformable or fragile objects requires not only structural compliance but also precise force-sensing capabilities. As illustrated in Figure 6a and Supplementary Video 3, FlexiRay, mounted on a UR5e robotic arm, sequentially grasps four soft and fragile objects from a table. Manually defined force thresholds are approximately 0.46 N for the bread (~20 g), 0.93 N for the chip (2.3 g), 1.12 N for the egg (44 g), and 1.88 N for the tape (25 g). These results demonstrate that FlexiRay achieves robust adaptability to delicate objects, ensuring grasping safety and stability through force closed-loop control, a capability critical for reliable operation in unstructured real-world grasping scenarios.

The capability to provide subtle tactile perception, coupled with inherent compliance, can be further leveraged in safer and more versatile human-robot interaction tasks. Figure 6b illustrates the actuation and control system for both the gripper's responsive evasion upon subtle contact and safe human-guided robotic arm manipulation, respectively. As shown in Figure 6c-d and Supplementary Video 4, upon detecting contact forces, the admittance controller generates compliant, force-adaptive motions based on pre-designed spring-damper characteristics, thereby ensuring interaction safety. However, it should be noted that large-area, low-texture objects (e.g., feathers) with minimal interaction forces pose challenges for conventional rigid vision-based tactile sensors, which rely heavily on the variation of surface normals. In

contrast, FlexiRay achieves fine-grained sensing by leveraging structural deformation in response to external loads and integrating proprioceptive information with contact region data. Figures 6c-d present the real-time force predictions collected during the experiments.

Fig. 6: Compliant grasping and safe human-robot interaction enabled by multi-modal perception.



a Compliant grasps of four soft or fragile objects, bread, chip, egg and tape, demonstrating force-closed-loop control and safe manipulation. **b** System architecture for responsive evasion upon subtle contact and human-guided robotic arm manipulation. **c** Subtle contact sensation via sensitive force sensing, exemplified by minute deformations induced by a feather. **d** Compliant response to contact using an admittance controller with tuned spring–damper characteristics, enabling safe, adaptive finger opening and closing in human–robot interaction. **e** Human–robot interface demonstration in which a single FlexiRay controls three degrees of freedom (DoFs) of a UR5e arm, mapping contact position and force directly to x, y and z motion in the world frame.

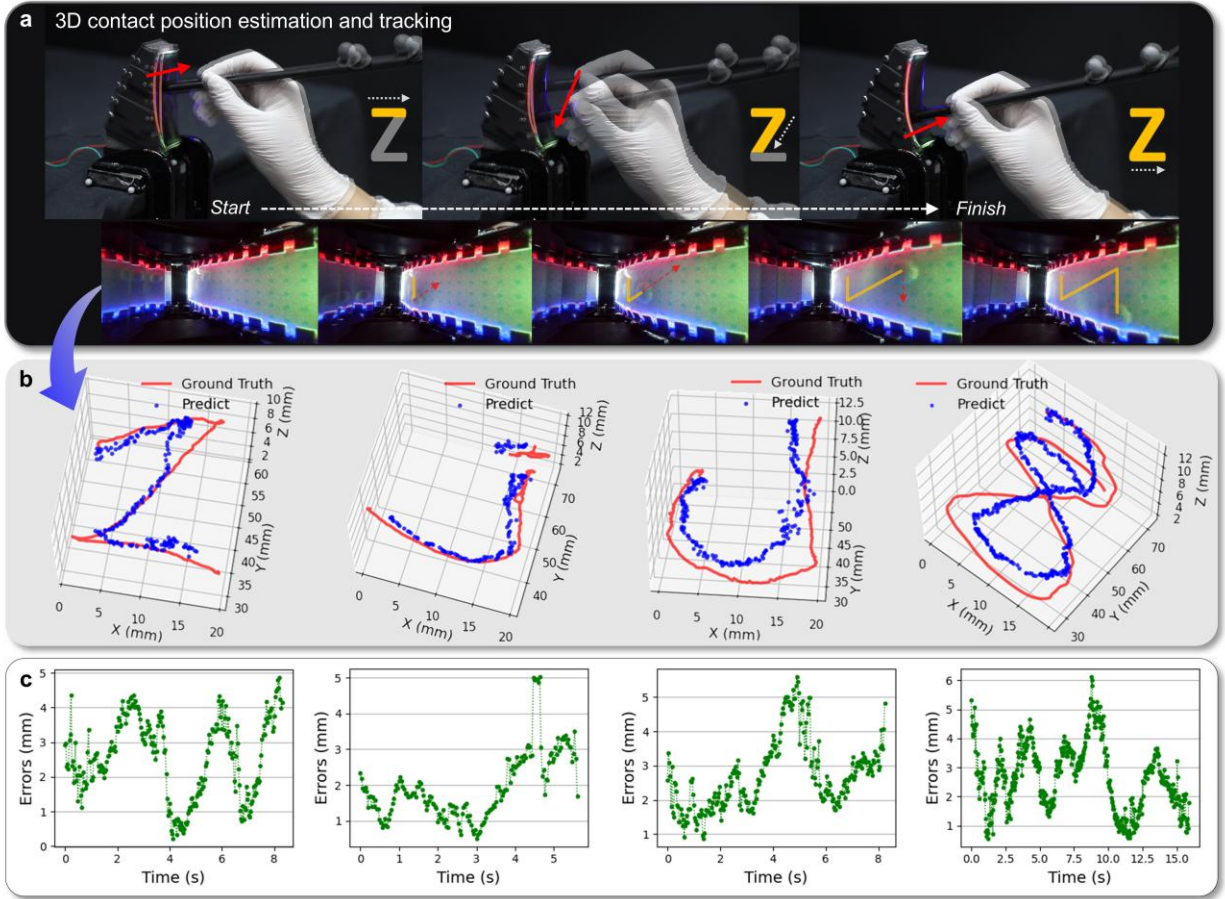
To further visualize the proposed model’s response mechanism across diverse contact scenarios, we employ the Structural Similarity Index (SSIM) and dense optical flow techniques to quantify continuous deformations and displacement directions of objects in sequential tactile images. As shown in Figure 6d, the multi-mirror array on the dorsal contact surface, combined with proprioception, amplifies subtle deformations during large-area contacts, ensuring that FlexiRay stably delivers accurate tactile perception under these challenging conditions. FlexiRay further demonstrate its potential as a human-robot interface, we show that a single FlexiRay can detect and control three degrees of freedom (DoFs) of a UR5 robotic arm based on its perception on contact position and force. In contrast to previous setups, the spring property of the admittance controller is set to 0. The initial contact position is defined as the interface origin, with the currently detected contact position and force used to control the UR5’s x, y, and z DoFs in the world coordinate system, respectively (see Fig. 6e and Supplementary Video 4).

Multi-modal tactile perception-enhanced intuitive physical human-robot interaction

Continuous contact localization during dynamic grasping is critical for stable grasping and human-robot interaction applications. To rigorously assess the contact localization capability of FlexiRay during dynamic interactions, we design a series of pen-tracing experiments using a high-precision visual motion capture system (NOKOV, calibrated positional accuracy <0.8 mm) as the ground-truth reference. In four independent trials, a human operator holds a pen-nib and inscribes distinct trajectories (the letters “Z,” “J,”

and “U,” as well as the numeral “8”) directly onto FlexiRay’s soft visuotactile sensing pad. Throughout these free-form motions, the contact localization model continuously estimates the interaction point in real time.

Fig. 7: Dynamic contact localization via free-form pen-tracing.



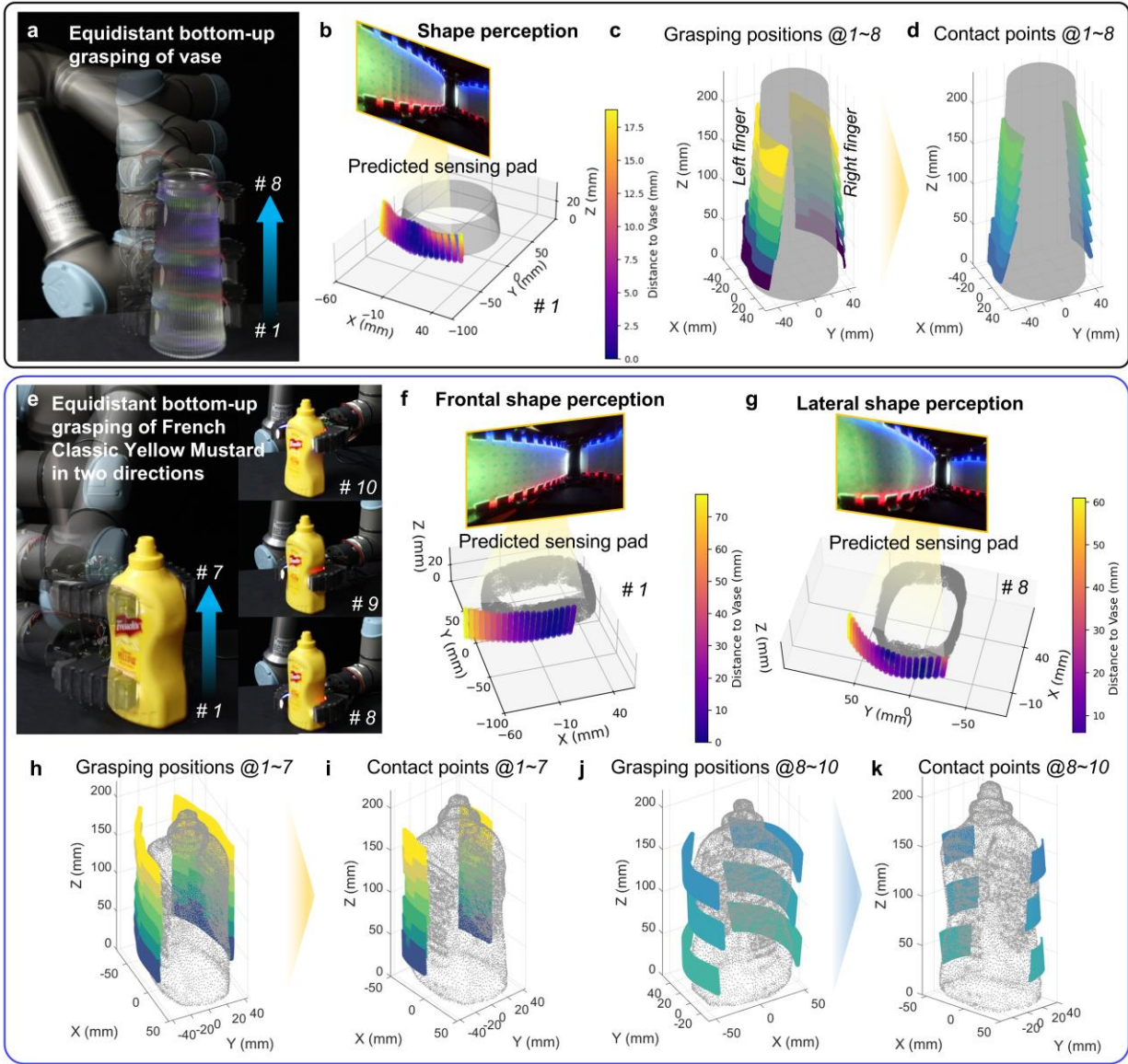
a Representative example of a “Z” trajectory drawn with a pen-nib directly onto FlexiRay’s soft visuotactile pad, with motion-capture ground truth and corresponding tactile images revealing spatial dynamics and evolving local deformation. **b** Predicted versus ground-truth trajectories for four free-form pen-traced characters (“Z,” “J,” “U,” and “8”). **c** Localization error distributions for contact points across the four trajectories, demonstrating consistently high spatial accuracy.

As shown in Fig. 7a, the writing of “Z” serves as a representative example, displaying the three-dimensional pen-tip trajectory alongside raw tactile images captured by the internal camera. This visualization captures not only the spatial dynamics of contact, but also the evolving local deformation and texture cues encoded during writing. Quantitative comparisons between predicted and ground-truth trajectories (Figs. 7b, c) reveal consistently high spatial fidelity: mean localization errors are 2.68 mm (standard deviation, s.d.: 1.15 mm) for the “8” trajectory, 2.73 mm (s.d.: 1.04 mm) for “U,” 2.58 mm (s.d.: 1.13 mm) for “Z,” and 1.93 mm (s.d.: 0.94 mm) for “J.” The relatively low s.d. values across trials confirm stable performance, even with variations in curvature, stroke direction, and contact dynamics. These results demonstrate that FlexiRay maintains sub-3-mm accuracy during non-repetitive, dynamic, multi-directional contact events—including under complex shear forces, structural compliance, and local material elasticity. The precision and robustness exhibited in this pen-tracing task highlight FlexiRay’s strong potential as an intuitive human–machine interface, capable of translating fine-grained contact position estimates into reliable control commands for robotic manipulation.

Global contact contour reconstruction

FlexiRay’s spatial proprioceptive perception further enables object shape reconstruction during dynamic grasping with improved efficiency. Two targets with distinct geometric characteristics were used, including a conical vase (truncated-cone lateral surface) and a French Classic Yellow Mustard container (low-curvature convex front surface, gourd-shaped side profile). Leveraging the high compliance, large contact coverage, and precise proprioceptive sensing, FlexiRay reconstructs contact contours efficiently and accurately with minimal trials (Supplementary Video 6).

Fig. 8: Shape perception of objects with distinct geometries using FlexiRay.



a Adaptive grasp sequence of the conical vase at eight evenly spaced heights from base to top. **b** Reconstructed tactile pad shapes at a single grasp position using proprioceptive sensing. Distances of individual points indicate grasp states and reveal local geometric features. **c** Reconstructed tactile pads from both fingers across all eight grasp heights. **d** Extracted local contact regions of the vase. **e** Shape perception of the French Classic Yellow Mustard container, including frontal and lateral grasps. Frontal tactile pad reconstruction at a single grasp (**f**) and across seven evenly spaced frontal positions (**h**), capturing the near-

planar, low-curvature front surface. Lateral tactile pad reconstruction at a single grasp (**g**) and across three lateral positions (**j**), revealing curvature variations along the gourd-shaped side profile. Local contact geometries from frontal (**i**) and lateral (**k**) grasps combined to reconstruct the container’s complete 3D contour.

For the vase, FlexiRay performed a series of adaptive grasps at evenly spaced heights from base to top, acquiring high-resolution tactile images at each stable contact (Fig. 8a). With its proprioceptive model, the spatial positions of key nodes on both fingers were estimated and mapped to a 3D coordinate frame. Spline interpolation was then applied to generate a continuous skeletal curve, followed by surface interpolation to reconstruct the tactile pads’ contact geometry (Fig. 8b). Contact regions extracted from the reconstructed surfaces revealed distinct local geometric features of the vase. Eight grasp heights were recorded (Fig. 8c), and integrating local information from different contact regions yielded a coherent analysis of the vase’s overall profile (Fig. 8d). Similarly, FlexiRay deployed adaptive grasps for the mustard container but across two distinct orientations. Seven positions on the front (Fig. 8h) captured the low-curvature near-planar geometry (Fig. 8f), while three lateral positions, aligned with different widths along the gourd-shaped profile (Fig. 8j), revealed its varying curvature (Fig. 8g). Local contact geometries from both orientations (Figs. 8i, k) were combined to further infer the object’s complete 3D contour.

These results demonstrate that FlexiRay efficiently acquires and integrates local shape features across multiple contact locations to generate a consistent object-level geometric representation. By combining compliant, adaptive grasping with large-area tactile imaging and accurate proprioceptive mapping, the system achieves precise surface contour perception without the need for time-consuming full-surface scanning. Such capabilities are particularly advantageous in unstructured scenarios where visual sensing is incomplete or occluded, offering both high perceptual efficiency and broad application potential.

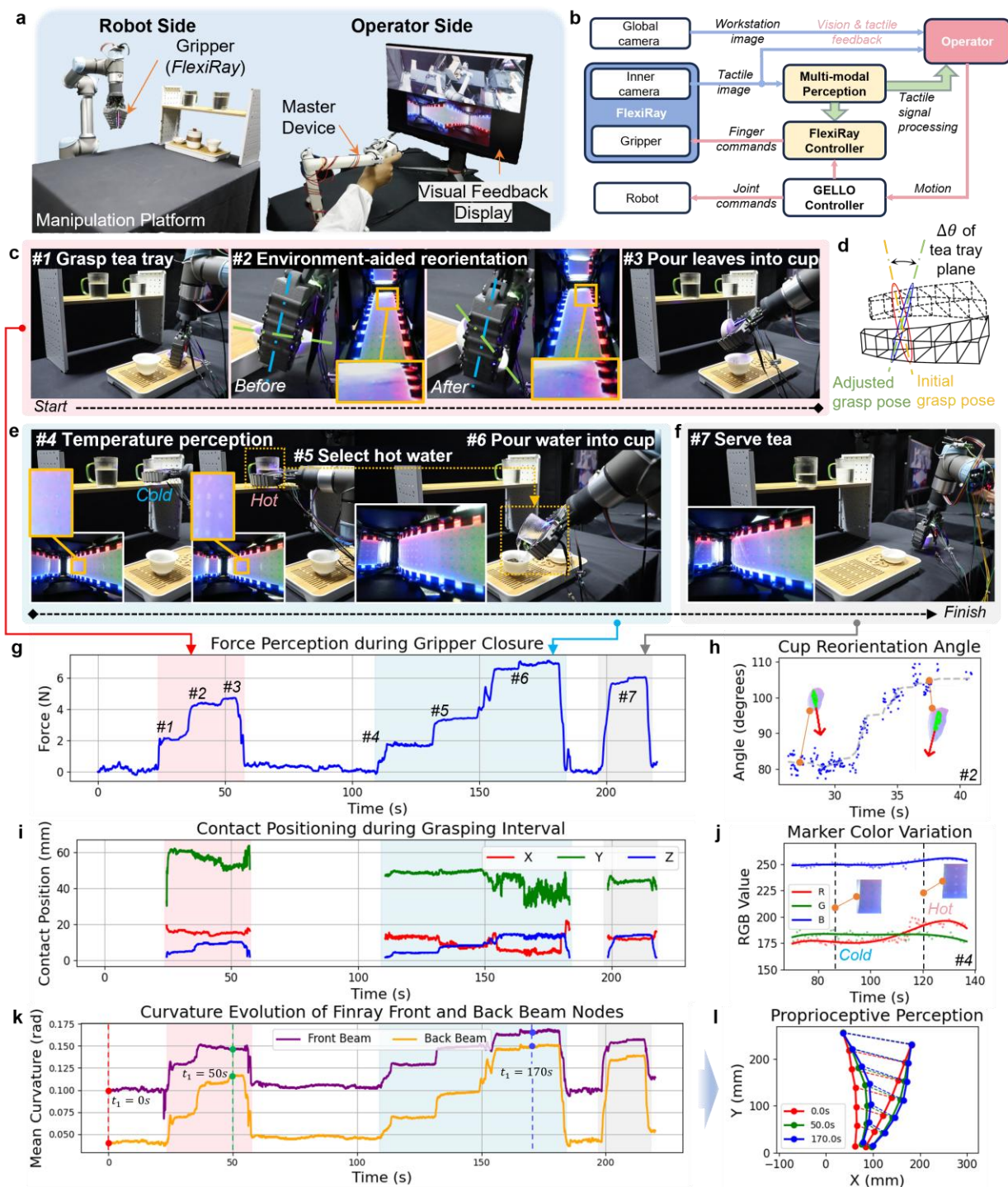
FlexiRay benchmarking via extended tea-preparation teleoperation

Currently, efficient learning of complex skills from human demonstrations has emerged as a key research direction for general robotic manipulation^{42–44}. However, existing studies predominantly focus on capturing only visual demonstration data, which limits dexterous manipulation in contact-rich environments and robots’ ability to interact with objects exhibiting different properties. In this work, we aim to demonstrate the potential of FlexiRay in providing richer multimodal information to enhance the quality of human demonstration data, using an extended-duration teleoperation task of traditional tea preparation as a case study.

This task integrates fine manipulation, in-hand tool reorientation, thermal discrimination, and stable pouring into a single continuous operation, which is not merely a procedural demonstration, but a systematical assessment that raises high requirements on the gripper’s multimodal sensing capabilities. Within an uninterrupted sequence, the system unifies force sensing, contact localization, proprioceptive deformation mapping, and temperature detection to underpin task execution. The tea-preparation task further acts as a controlled yet challenging platform for validating such multimodal perception. It demands the operator adapt to unpredictable object dynamics (e.g., loose tea leaf movement, liquid sloshing), discriminate visually similar objects via temperature, and execute precise pouring, all under remote teleoperation, where direct physical contact is absent and visual occlusion occurs frequently. In this scenario, FlexiRay redefines conventional vision- and force-dominated teleoperation, transforming it into a multisensory manipulation process. By leveraging its inherent compliance and safe physical interaction, the gripper enables operators to make real-time decisions at every stage, supported by continuous feedback from its integrated sensing modalities. Specifically it requires the operator to adapt to unpredictable object dynamics (e.g., loose tea leaf movement, liquid sloshing), discriminate visually similar objects via temperature, and perform precise pouring, all under remote teleoperation, where direct physical contact is absent and visual occlusion occurs frequently.

In the reorientation stage, FlexiRay leverages precise contact localization and force feedback, allowing the operators to adjust the teapot tray’s pose without releasing it. For instance, the initial grasp positioned the tray plane at an $\sim 80^\circ$ offset relative to the gripper’s z-axis, which is suboptimal for pouring. Conventional teleoperation, due to visual occlusion or the lack of accurate estimation of in-hand object pose and contact distribution, necessitates large, unstable arm movements or complete regrasping. In contrast, FlexiRay enables controlled micro-sliding against external supports, with visibility and continuous monitoring throughout. Real-time tracking of contact texture orientation and contact point trajectories prevents slippage, facilitating a smooth 25° in-hand reorientation (Figs. 9c, j) and allowing pouring from a mechanically advantageous arm configuration. In the thermal identification step, FlexiRay’s temperature-sensitive tactile pads provide instantaneous feedback via colorimetric markers (Fig. 9e), enabling operators to distinguish between room-temperature and hot water cups without repeated confirmation. The chromaticity change curve (Fig. 9j) shows clear separation between “cold” and “hot” contact intervals, underscoring the thermal channel’s high sensitivity—a capability rarely matched by conventional teleoperation systems. Notably, FlexiRay’s proprioceptive sensing stably predicts and visualizes the deformation of its compliant beams throughout operations, delivering real-time data on interaction dynamics and grasp quality and remains fully operational even under visual occlusion. Figures 9g–k depict force, contact position, and proprioceptive signals (encoded as curvature of anterior and posterior beam nodes) across the manipulation sequence. Larger curvature values correspond to stronger grip forces, consistent with the passive deformation properties of the FinRay structure. Figure 9l illustrates predicted gripper deformation at three key moments (0 s, 50 s, 170 s), offering intuitive guidance for refining grasp strategies.

466 **Fig. 9: Extended teleoperation with FlexiRay for remote tea preparation.**



467

468 **a** Experimental platform for the teleoperated tea preparation task, comprising the robot side and the operator

469 side. **b** Control architecture of the teleoperation system, integrating a GELLO-based teleoperation

controller⁴⁵ with FlexiRay’s multimodal tactile perception capabilities—force feedback, contact localization, proprioception, and temperature sensing. **c** Sequential task steps for teleoperated tea preparation, where real-time tactile feedback from FlexiRay is displayed to the operator, ensuring stable manipulation despite visual occlusion and unpredictable liquid dynamics. Steps #1–3 involve in-hand reorientation of the teapot tray via controlled micro-sliding against an external support. **d** Geometric model representation of the in-hand tray reorientation, with **(h)** showing the corresponding real-time computation of deflection angles. **e** Thermal discrimination between hot and room-temperature cups via the temperature-sensitive tactile pad, with color change used to identify hot water. **j** presents the RGB marker color variations. **f** Steps of tea pouring and cup delivery. Throughout the operation, multimodal signals are recorded in real time, including force **(g)**, contact position **(i)**, and curvature of anterior and posterior compliant beams for proprioception **(k)**. **l** Predicted gripper deformation at three key moments (0 s, 50 s, 170 s), providing intuitive feedback for grasp strategy optimization.

Long-horizon teleoperation experiments clearly demonstrate FlexiRay’s multimodal perception capabilities, which integrate all five sensory modalities to assist human operators in achieving more precise and safer manipulation. Even under conditions of visual occlusion or partial visual information loss, FlexiRay enables reliable task execution, highlighting its potential to enhance robotic systems in adapting to dynamic and complex real-world environments. These experiments substantiate the practical advantages of multimodal sensing, showing how the fusion of force, temperature, texture, contact localization, and proprioception leads to highly adaptive and efficient robotic systems. Looking forward, integrating FlexiRay with imitation learning frameworks is expected to enable safer and higher-quality transfer of human demonstrations to robotic systems, paving the way toward more generalizable robotic manipulation.

Discussion

This study presents a novel VTS-integrated flexible robotic gripper, Tactile-integrated FlexiRay, inspired by human tactile modalities and the FinRay effect. The gripper effectively combines enhanced structural compliance with advanced multimodal sensing capabilities, successfully realizing five out of the seven primary human tactile modalities, including contact force, location, texture, temperature, and proprioception, excluding pain and vibration. Specifically, the normal force estimation accuracy reaches

0.17 N, the contact location accuracy is 0.96 mm, and the spatial proprioception accuracy is 0.24 mm.

Through strategic structural design and optimization, FlexiRay demonstrates strong resistance to visual interference even under large deformations. The average effective sensing coverage of the tactile sensing pad across different deformation states is 87.2%, with over 70% maintained during large deformations. To address the visual occlusion gaps that occur during large deformations, we do not limit the compliance of the finger to maintain the optical path. Instead, we treat deformation-induced optical path disturbances as key design parameters. We innovatively design a segmented mirror array and proposes a method for optimizing the internal optical system layout. We characterize the force-deformation patterns of the FRE fingers under different loading conditions in a physical environment. These patterns serve as the basis for determining the optimal position and orientation of the mirrors. This method ensures consistent image acquisition through passively controlled mirror reflections, even under substantial deformation. Therefore, FlexiRay excels in grasping irregular, cylindrical, or spherical objects, significantly increasing the compliant contact area and enhancing the richness and efficiency of the perception data during flexible interactions.

Compared to existing VTS technologies, FlexiRay demonstrates a pronounced compliance advantage, directly enhancing grasping universality, perceptual robustness, and interaction safety. Unlike rigid VTS designs that require precise alignment or additional actuation before acquiring valid tactile information, FlexiRay achieves secure and gentle handling of fragile or irregular objects solely through passive adaptation. Its compliant structure maintains stable large-area tactile perception even under substantial

519 deformation, effectively overcoming the coverage loss commonly observed in rigid systems due to visual
520 occlusion. Furthermore, the Fin Ray–based compliance provides inherent safety in human–robot interaction:
521 when excessive forces occur, the structure absorbs the overload through large deformations, preventing
522 damage to both the object and the sensor.

523 Among existing flexible VTS designs, the most comparable is the GelSight Baby Fin Ray, yet FlexiRay
524 exhibits a much stronger compliance and perceptual capability. Reported data show that under a contact
525 force of 7.5 N, the Baby Fin Ray deforms only 3–4 mm, whereas FlexiRay achieves a deformation of ~15
526 mm in the contact depth direction (over four times greater). Moreover, FlexiRay retains an effective sensing
527 area above 1092 mm² during large deformations, with an average of ~1360 mm² across deformation states.
528 This level of stability is not observed in current VTS systems. The combination of high compliance and
529 broad perceptual coverage allows FlexiRay to provide conformal, stable grasping while capturing richer
530 tactile information from a single interaction.

531 Building on these advantages, FlexiRay achieves 96.7% success in a texture recognition task with
532 perception balls, surpassing the 86.7% success rate of GelSight Mini. In addition, through multimodal
533 sensing and a deep learning model, FlexiRay attains proprioceptive capabilities analogous to the human
534 tactile system: in human–robot interaction, it demonstrates sensitive force perception and reliable contact
535 localization, while in object shape reconstruction, it confirms robust sensing of its proprioceptive
536 deformation states. Furthermore, within the developed teleoperation framework, FlexiRay delivers real-
537 time multimodal feedback—including temperature perception—extending the operator’s perceptual
538 boundary during long-horizon and fine-grained tasks. These results underscore its unique potential for
539 enhancing safety and human–robot collaboration.

540 In terms of cost and availability, the affordability of the imaging components and the carefully engineered
541 fabrication process provide FlexiRay with a distinct advantage in both manufacturing cost and processing
542 complexity. Unlike approaches such as GelSight, which require strong interlayer adhesion of coatings,
543 FlexiRay can be readily produced by integrating components and casting materials onto a 3D-printed TPU

beam framework, enabling seamless monolithic molding. During data acquisition, the same FlexiRay endured more than 40,000 repeated interactions while maintaining stability and accuracy, demonstrating its durability. The modular design of the anterior and lateral beams allows the tactile sensing pads to be easily replaced, simplifying maintenance and extending device lifespan.

Overall, FlexiRay represents a significant advancement in the development of high-resolution, large-coverage, and cost-effective perception capabilities for soft robotic systems. Future research will focus on integrating adjustable skeletal stiffness and advanced perception field materials to further enhance the realism of perception, particularly in texture sensing. Additionally, we plan to expand FlexiRay's capabilities to multi-fingered grippers and explore its applications in dynamic, high-safety tasks such as fruit picking and assistive care, paving the way for the next generation of robust, adaptable, and intelligent human-robot interaction systems.

Methods

Optical system implementation

- **Camera.** A camera with high-resolution image acquisition capability (12 million pixels) is selected, featuring a compact form factor ($8 \times 8.5 \times 5$ mm) and a wide-angle field of view (135°). Its compact form enables integration into the gripper, capturing high-clarity internal imagery with sufficient detail of contact interfaces.
- **Illumination.** Flexible RGBW (red, green, blue, white) LED lights are embedded within the silicone gel in a series-connected configuration during the fabrication process. To address the issue of ambient light interference, the system incorporates interlaced side beams. This beam design not only effectively mitigates interference from external light sources but also retains the flexibility required for compliant and adaptive interactions.

- Mirrors. The mirrors are strategically positioned and mounted at a specific angular orientation on the finger's skeletal structure to extend the visual field of view. This configuration ensures comprehensive visual coverage of the interaction area even when the finger undergoes significant deformation during operation. To preserve the compliance of the finger structure without introducing structural constraints, the rigid planar mirror adopts a T-shaped design. This geometric configuration minimizes the bonding area with the flexible back beam of the finger, thereby reducing mechanical interference with the beam's deformation. Simultaneously, the T-shaped structure ensures that the effective reflective surface area remains sufficiently large to maintain the required optical path efficiency.

Skeleton and tactile sensing pad implementation

- Finger. The Fin Ray finger, fabricated from TPU materials, exhibits intrinsic compliance that enables it to passively conform to the contours of various objects without relying on active actuation mechanisms.
- Tactile Sensing Pad. The tactile sensing pad primarily consists of five functional layers. The light-shielding layer is fabricated by blending Smooth-On Inc. 00-30 silicone with black color paste, aiming to block the interface from environmental light. The temperature-sensing layer is formulated by incorporating three thermochromic pigments into the silicone, leveraging the pigments' reversible color-changing properties in response to temperature variations to enable non-intrusive thermal detection. For the reflective layer, silicone is mixed with aluminum powder and aluminum flake in a mass ratio of 400 : 20 : 3. The elastic layer utilizes a low-hardness transparent silicone (Shore 0A), whose ultra-soft mechanical properties facilitate precise transduction of surface texture information through deformations. Additionally, polydimethylsiloxane (PDMS) is employed as an elastic support layer, designed to maintain the overall compliance of the sensing pad while providing structural reinforcement to preserve deformation capacity during interactions with objects.

Fabrication and manufacturing

Our primary objective is to endow FRE with tactile perception capabilities through a vision-based tactile sensing method while preserving its inherent compliance and passive deformation characteristics. To achieve this, the conventional acrylic support structure is replaced with PDMS material, and flexible LED strips are adopted for illumination. The fabrication process is detailed as follows:

First, a reflective layer is fabricated by mixing Smooth-On Inc. 00-30 silicone with aluminum powder and flake according to the specified mass ratio. To ensure surface quality, the mixture is poured onto a glass substrate, spread to form a thin film (thickness: 0.25 mm) with a squeegee applicator, and cured at 80 °C for 30 minutes. Subsequently, an array of round markers is removed from the cured film using a laser cutter. Next, a temperature-sensing layer is prepared by mixing the silicone with three thermochromic pigments in a mass ratio of 20:1:1:1. These pigments exhibit distinct thermochromic responses: one transitions from colorless to blue at 18 °C, another from green to yellow at 38 °C, and the third from colorless to magenta at 60 °C. Consequently, this combination enables a temperature-dependent color transition of the thermometric layer: it appears pink above 60 °C, yellow between 38 °C and 60 °C, deep blue below 18 °C, and dark green under ambient conditions. Using the same application method, a 0.15 mm thick temperature-sensing layer is produced and bonded to the reflective layer. The temperature-dependent color of the thermo-sensitive silicone is shown in Supplementary Fig. S7. To avoid the interference from external ambient light, a 0.15-millimeter-thick black light-shielding layer is applied by mixing silicone with black pigment paste and coating onto the previous layer.

The prepared composite layer is then removed and placed into a mold. A low-hardness, transparent silicone (mixed in a ratio of 1.8:1) is poured over it to create a soft, elastic layer with a Shore A hardness of 0, approximating that of human skin. To ensure a uniformly smooth surface, the material is smoothed out with a squeegee and cured at room temperature for 90 minutes. Subsequently, the RGBW LEDs are connected in series and affixed to the TPU front beam. PDMS is mixed in a ratio of 5:1 to form a transparent support structure capable of withstanding large deformations. The TPU front beam with integrated LEDs is

positioned within the mold. Further, PDMS is poured to encapsulate these components, resulting in a tactile sensing pad capable of perceiving texture, temperature, and other contact information.

Following the fabrication of the tactile sensing silicone pad, the assembly of the FlexiRay is completed by integrating the mirrors, internal camera, side beams, finger framework, and the soft sensing pad. To validate its perception and grasping capabilities, a gripper is designed comprising two identical FRE fingers. A single stepper motor actuates the gripper through linkage mechanisms, enabling a substantial range of opening and closing motions.

CMA-ES based optics layout optimization

The equilibrium between finger compliance and the robustness of the tactile sensing region presents a significant challenge in the advancement of soft VTS. To mitigate the limitations of FOV and occlusion blind spots associated with large-deformation finger structures, a multi-mirror FOV optimization method is proposed, leveraging CMA-ES^{46,47}. Through the optimization of both structural and layout parameters of the camera and mirrors, a visual reflection system composed of multiple planar mirrors is established on the back side of the finger. This configuration facilitates a single camera in achieving comprehensive coverage of the entire deformation range and contact region of the Fin Ray, thereby enabling enhanced visual-tactile perception.

Initially, the displacement of nodes under the applied force F is collected for the Fin Ray, which has identical structural parameters. The displacement is represented by the coordinates of the lateral 2D cross-section, with the back joint nodes denoted as $\{\mathbf{N}_i\}_{i=1}^n$ and the tactile sensing area joint nodes labeled as $\{\mathbf{P}_i\}_{i=1}^n$. Consequently, a mapping from force to deformation is established, represented as $f: F \rightarrow \{\mathbf{N}_i, \mathbf{P}_i\}_{i=1}^n$. The subsequent part provides a detailed account of the construction and solution of the optical layout optimization problem.

Decision variables:

$$\mathbf{x} = \{(\theta_i^{\text{mir}}, t_i^{\text{mir}}, l_i^{\text{mir}})\}_{i=1}^{n-1} \cup \{u, \phi\}, \quad \text{with } \theta_i^{\text{mir}}, t_i^{\text{mir}}, l_i^{\text{mir}}, u, \phi \in \mathbb{R} \quad (1)$$

where $\{\theta_i^{\text{mir}}, t_i^{\text{mir}}, l_i^{\text{mir}}\}$ represents the parameters associated with the mirrors. Specifically, θ_i^{mir} denotes the angle between the mirror and the line segment formed by two adjacent nodes on the back side, t_i^{mir} indicates the midpoint offset distance, and l_i^{mir} signifies the length of the mirror. The camera is positioned along the baseline defined by the back side and the base nodes of the tactile sensing pad. The distance from the camera to the back base node is expressed as a coefficient u , relative to the length of the baseline. Furthermore, the angle between the optical axis of the camera and the baseline is represented by ϕ .

Objective function:

Given a camera with a fixed FOV, the discrete rays produced by the camera can be represented as $\mathcal{R} = \{\mathbf{r}_j\}_{j=1}^m$. The primary objective of the optimization process is to maximize the coverage of the tactile sensing region by the collection of FOV rays across various deformations. This goal can be achieved through two approaches: direct imaging and single reflections from mirrors.

By sampling K distinct loads from the force-displacement mapping f , a set of joint nodes \mathcal{D} is obtained. Each deformation structure \mathbf{d} corresponds to a discrete set of target points \mathbf{p} , denoted as $\mathcal{P}^{\mathbf{d}}$, within the tactile sensing region. The radius of the illumination range for a single FOV ray is defined as $R \propto l^c$, which depends on the propagation distance l^c . Consequently, the objective function can be expressed as follows:

$$\text{Maximize } f(\mathbf{x}) = \frac{1}{K} \sum_{\mathbf{d} \in \mathcal{D}} \sum_{\mathbf{r} \in \mathcal{R}} \sum_{\mathbf{p} \in \mathcal{P}^{\mathbf{d}}} I(\mathbf{x}, \mathbf{r}, \mathbf{p}, R) \quad (2)$$

where $I(\mathbf{x}, \mathbf{r}, \mathbf{p}, R)$ is an indicator function that returns 1 if the target point resides within the coverage range of the ray beam; otherwise, it returns 0.

Constraints:

The following constraints must be adhered to:

- **Geometric constraints.** The lengths, offsets, and rotation angles of each mirror, as well as the position of the camera, must remain within predefined ranges: $l_i^{\text{mir}} \in [l_{\min}, l_{\max}]$, $t_i^{\text{mir}} \in [t_{\min}, t_{\max}]$, $\theta_i^{\text{mir}} \in [\theta_{\min}, \theta_{\max}]$, $u \in (0,1)$.
- **Occlusion constraints.** The potential occlusion between mirrors during the deformation process must be taken into account. If an occlusion exists along the line of sight, the indicator function $I(\mathbf{x}, \mathbf{r}, \mathbf{p}, R)$ returns 0.
- **Safety constraints.** Under all deformation conditions, mirrors must not interfere with the front and back beams. If this condition is not met, the indicator function $I(\mathbf{x}, \mathbf{r}, \mathbf{p}, R)$ returns 0 as a penalty.

Solution process:

The optimization process utilizes the CMA-ES algorithm to search for optimal layout solutions by sampling candidate solutions from a multivariate Gaussian distribution. The multivariate Gaussian distribution is defined as $\mathcal{N}(\boldsymbol{\mu}, \sigma^2 \mathbf{C})$, where $\boldsymbol{\mu}$ is the mean vector, \mathbf{C} is the covariance matrix, and σ represents the step size.

In the y -th generation of the optimization process, for a population size of β , the candidate solutions $\{\mathbf{x}_i\}_{i=1}^{\beta}$ are obtained by sampling from the distribution $\mathcal{N}(\boldsymbol{\mu}^{(y)}, (\sigma^{(y)})^2 \mathbf{C}^{(y)})$:

$$\mathbf{x}_i = \boldsymbol{\mu}^{(y)} + \sigma^{(y)} \sqrt{\mathbf{C}^{(y)}} \mathbf{s}_i \quad (3)$$

where $\mathbf{s}_i \sim \mathcal{N}(0, \mathbf{I})$, with \mathbf{I} denoting the identity matrix. The k solutions with the highest objective function values are selected as candidates: $\{f(\mathbf{x}_{1:k}) \mid f(\mathbf{x}_1) \geq f(\mathbf{x}_2) \geq \dots \geq f(\mathbf{x}_k) \geq \dots \geq f(\mathbf{x}_{\beta})\}$

The evolution paths are updated using the following equations:

$$\mathbf{p}_{\sigma}^{(y+1)} = (1 - c_{\sigma}) \mathbf{p}_{\sigma}^{(y)} + \sqrt{c_{\sigma}(2 - c_{\sigma}) \lambda_w} \sqrt{\mathbf{C}^{(y)}}^{-1} \quad (4)$$

$$\mathbf{p}_c^{(y+1)} = (1 - c_c) \mathbf{p}_c^{(y)} + \sqrt{c_c(2 - c_c) \lambda_w} H_{\sigma}^{(y+1)} \quad (5)$$

679 where c_σ and c_c are cumulation factors, respectively. $H_\sigma^{(y+1)}$ is the Heaviside function, and

$$680 \quad \delta = \sum_{i=1}^{\eta} w_i \sqrt{\mathbf{C}^{(y)}} \mathbf{s}_i, \quad \lambda_w = \frac{1}{\sum_{i=1}^{\eta} w_i^2}, \quad \sum_{i=1}^{\eta} w_i = 1$$

681 Therefore, the parameters of the multivariate Gaussian distribution for each generation are updated as
682 follows:

$$683 \quad \boldsymbol{\mu}^{(y+1)} = \boldsymbol{\mu}^{(y)} + c_l \sigma^{(y)} \sqrt{\mathbf{C}^{(y)}} \mathbf{s}_i \quad (6)$$

$$684 \quad \sigma^{(y+1)} = \sigma^{(y)} \exp \left(1, c_\sigma \left(\frac{\|\mathbf{p}_\sigma^{(y+1)}\|}{\mathbb{E}[\mathcal{N}(0, \mathbf{I})]} - 1 \right) / d_\sigma \right) \quad (7)$$

$$685 \quad \mathbf{C}^{(y+1)} = \left(1 + c_1 c_\mu (2 - c_c) \left(1 - H_\sigma^{(y+1)} \right) \right) \mathbf{C}^{(y)} + c_1 \Delta_1 + c_\mu \Delta_\mu \quad (8)$$

686 where Δ_1 and Δ_μ correspondingly denote the rank one and $-\mu$ updates. The parameters c_l , c_1 , and c_μ
687 represent the learning rates for the mean, rank one, and $-\mu$ updates, respectively. The damping factor d_σ
688 is used for the adaptive accumulation of step sizes.

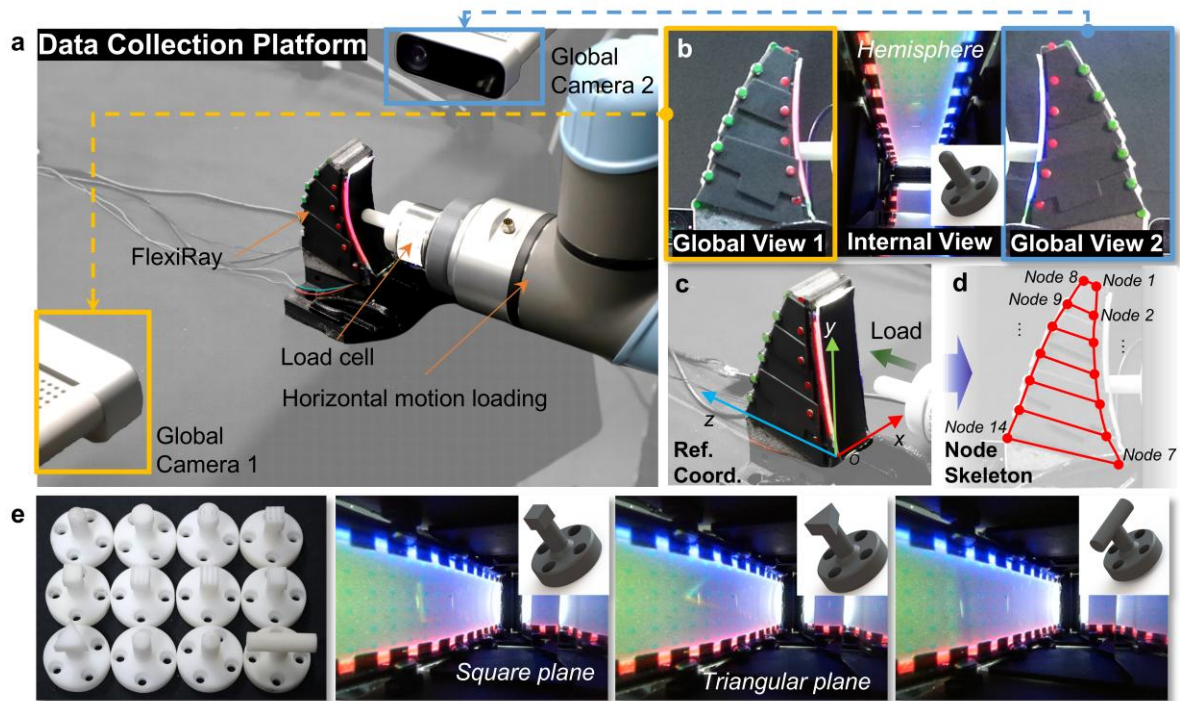
689 Starting with initial parameters that satisfy the constraints, the parameters of the multivariate Gaussian
690 distribution are updated iteratively. This process continues until either the expected stopping condition for
691 the objective function is met or the maximum number of optimization generations is reached. The optimized
692 camera and mirror layout parameters are then derived from the distribution. [Sensitivity analysis of the](#)
693 [decision variables and hyperparameter analysis of the optimization algorithm are detailed in the](#)
694 [supplementary material \(Figs. S1 and S2\).](#)

695 Proprioception data collection platform and procedures

696 The dataset collection platform for proprioception tasks is shown in Fig. 10a. FlexiRay is mounted at a
697 fixed position on an optical bench, with the reference coordinate system established at the lower-left corner
698 of the tactile pad (Fig. 10c). A six-axis force sensor is attached to the end of a UR5e robotic arm, whose

tool coordinate system is aligned such that its z-axis is parallel to the z-axis of the reference frame, enabling normal force application. A 3D-printed contact probe is mounted on the sensor, designed to simulate diverse contact conditions through 12 interchangeable geometries—including hemispherical, square, triangular, and cylindrical tips (Fig. 10e, full set in Supplementary Fig. S5). During data collection, the probe is randomly positioned within the feasible contact domain in the x–y plane, randomly rotated around the z-axis, and then advanced along the z-axis until contact is detected by the force sensor, which triggers synchronized acquisition. At this stage, three data streams are recorded simultaneously: lateral beam deformations captured by dual global cameras (Fig. 10b), tactile images from the internal camera, and force readings from the load cell. In parallel, proprioceptive node trajectories are extracted in real time (Fig. 10d), while the robotic end-effector position is continuously logged to ensure precise temporal and spatial alignment across all modalities.

Fig. 10: Data collection platform and procedures.



a Dataset collection platform for proprioception tasks. **b** Deformation representation via planar views of FlexiRay's two lateral beams captured by dual global cameras, alongside raw tactile images recorded by the internal camera. **c** Reference coordinate system for contact localization, established at the lower-left corner of the tactile pad. **d** Extracted skeletal structure of proprioceptive nodes. **e** Probe types attached to the load cell, including hemispherical, square, triangular, and cylindrical tips, among a total of 12 geometries (full set provided in Supplementary Fig. S5).

Machine learning implementation

In terms of model implementation, the PP-LiteSeg semantic segmentation model is built upon the STDC2 backbone, with the objective function using Cross-Entropy Loss. The model is optimized using Stochastic Gradient Descent (SGD) with momentum set to 0.9 and weight decay of 4×10^{-5} . A polynomial decay learning rate scheduler is employed, with an initial learning rate of 0.01 and a decay factor of 0.9. After training, the model's performance evaluation shows a mean Intersection over Union (mIoU) of 89.98%. For the proprioception model, three heads are used to predict normal contact force, contact position, and node localization, each consisting of fully connected layers with 512 and 256 neurons. The output layers have dimensions of \mathbb{R}^1 , \mathbb{R}^3 and \mathbb{R}^{28} , respectively. The loss function for all three outputs is Mean Squared Error (MSE), with the learning rate set to 0.001. The dataset comprises 5,000 tactile image samples, which are split into training and validation sets at a 4:1 ratio. Additionally, the classification head of the texture recognition model is composed of fully connected layers with 512 and 256 neurons. The loss function used is Cross-Entropy Loss, and the model is trained with the Adam optimizer, with a learning rate set to 0.001.

Data Availability

The data that support the findings of this study are available within the paper and the Supplementary data files. Other data generated during the current study are available from the corresponding author on reasonable request.

Code Availability

The demo implementation of proprioception model is provided in the Supplementary Dataset files.

References

1. Dahiya, R. S., Metta, G., Valle, M. & Sandini, G. Tactile sensing—from humans to humanoids. *IEEE transactions on robotics* **26**, 1–20 (2009).
2. Felten, D. L., O'Banion, M. K. & Maida, M. S. 9 - Peripheral Nervous System. in *Netter's Atlas of Neuroscience (Third Edition)* (eds. Felten, D. L., O'Banion, M. K. & Maida, M. S.) 153–231 (Elsevier, Philadelphia, 2016).
3. Handler, A. & Ginty, D. D. The mechanosensory neurons of touch and their mechanisms of activation. *Nature Reviews Neuroscience* **22**, 521–537 (2021).
4. She, Y., Liu, S. Q., Yu, P. & Adelson, E. Exoskeleton-covered soft finger with vision-based proprioception and tactile sensing. in *2020 IEEE International Conference on Robotics and Automation (ICRA)* 10075–10081 (IEEE, 2020).
5. Liu, Z. *et al.* A three-dimensionally architected electronic skin mimicking human mechanosensation. *Science* **384**, 987–994 (2024).
6. Zhou, Z. *et al.* A sensory soft robotic gripper capable of learning-based object recognition and force-controlled grasping. *IEEE Transactions on Automation Science and Engineering* **21**, 844–854 (2022).
7. Yan, Y. *et al.* Soft magnetic skin for super-resolution tactile sensing with force self-decoupling. *Science Robotics* **6**, eabc8801 (2021).
8. Taylor, S., Park, K., Yamsani, S. & Kim, J. Fully 3D printable Robot Hand and Soft Tactile Sensor based on Air-pressure and Capacitive Proximity Sensing. in *2024 IEEE International Conference on Robotics and Automation (ICRA)* 18100–18105 (2024).
9. Shorthose, O., Albini, A., He, L. & Maiolino, P. Design of a 3D-Printed Soft Robotic Hand With Integrated Distributed Tactile Sensing. *IEEE Robotics and Automation Letters* **7**, 3945–3952 (2022).
10. Zhao, H., O'Brien, K., Li, S. & Shepherd, R. F. Optoelectronically innervated soft prosthetic hand via stretchable optical waveguides. *Science Robotics* **1**, eaai7529 (2016).
11. Meerbeek, I. M. V., Sa, C. M. D. & Shepherd, R. F. Soft optoelectronic sensory foams with proprioception. *Science Robotics* **3**, eaau2489 (2018).
12. Wall, V., Zöller, G. & Brock, O. Passive and active acoustic sensing for soft pneumatic actuators. *The International Journal of Robotics Research* **42**, 108–122 (2023).
13. Li, G., Liu, S., Wang, L. & Zhu, R. Skin-inspired quadruple tactile sensors integrated on a robot hand enable object recognition. *Science Robotics* **5**, eabc8134 (2020).
14. Spiegel, J. V. der. Computational sensors: The basis for truly intelligent machines. in *Intelligent Sensors* (ed. Yamasaki, H.) vol. 3 19–37 (Elsevier Science B.V., 1996).
15. Abad, A. C., Reid, D. & Ranasinghe, A. HaptiTemp: A next-generation thermosensitive GelSight-like visuotactile sensor. *IEEE Sensors Journal* **22**, 2722–2734 (2021).

16. Fang, B. *et al.* Force Measurement Technology of Vision-Based Tactile Sensor. *Advanced Intelligent Systems* 2400290 (2024).
17. Xu, W. *et al.* A Vision-Based Tactile Sensing System for Multimodal Contact Information Perception via Neural Network. *IEEE Transactions on Instrumentation and Measurement* (2024).
18. Abad, A. C. & Ranasinghe, A. Visuotactile sensors with emphasis on gelsight sensor: A review. *IEEE Sensors Journal* **20**, 7628–7638 (2020).
19. Johnson, M. K. & Adelson, E. H. Retrographic sensing for the measurement of surface texture and shape. in *2009 IEEE Conference on Computer Vision and Pattern Recognition* 1070–1077 (2009).
20. Kakani, V., Cui, X., Ma, M. & Kim, H. Vision-based tactile sensor mechanism for the estimation of contact position and force distribution using deep learning. *Sensors* **21**, 1920 (2021).
21. Yuan, W., Dong, S. & Adelson, E. H. Gelsight: High-resolution robot tactile sensors for estimating geometry and force. *Sensors* **17**, 2762 (2017).
22. Yuan, W., Mo, Y., Wang, S. & Adelson, E. H. Active clothing material perception using tactile sensing and deep learning. in *2018 IEEE International Conference on Robotics and Automation (ICRA)* 4842–4849 (IEEE, 2018).
23. Li, S. *et al.* M³Tac: A Multispectral Multimodal Visuotactile Sensor With Beyond-Human Sensory Capabilities. *IEEE Transactions on Robotics* **40**, 4506–4525 (2024).
24. Do, W. K. & Kennedy, M. Densetact: Optical tactile sensor for dense shape reconstruction. in *2022 International Conference on Robotics and Automation (ICRA)* 6188–6194 (IEEE, 2022).
25. Tippur, M. H. & Adelson, E. H. GelSight360: An omnidirectional camera-based tactile sensor for dexterous robotic manipulation. in *2023 IEEE International Conference on Soft Robotics (RoboSoft)* 1–8 (IEEE, 2023).
26. Lambeta, M. *et al.* Digit: A novel design for a low-cost compact high-resolution tactile sensor with application to in-hand manipulation. *IEEE Robotics and Automation Letters* **5**, 3838–3845 (2020).
27. Donlon, E. *et al.* Gelslim: A high-resolution, compact, robust, and calibrated tactile-sensing finger. in *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)* 1927–1934 (IEEE, 2018).
28. Lepora, N. F., Lin, Y., Money-Coomes, B. & Lloyd, J. DigiTac: A DIGIT-TacTip Hybrid Tactile Sensor for Comparing Low-Cost High-Resolution Robot Touch. *IEEE Robotics and Automation Letters* **7**, 9382–9388 (2022).
29. Zhang, X., Yang, T., Zhang, D. & Lepora, N. F. Tacpalm: A soft gripper with a biomimetic optical tactile palm for stable precise grasping. *IEEE Sensors Journal* (2024).
30. Gomes, D. F., Lin, Z. & Luo, S. GelTip: A Finger-shaped Optical Tactile Sensor for Robotic Manipulation. in *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)* 9903–9909 (2020). doi:10.1109/IROS45743.2020.9340881.
31. Zhao, J. & Adelson, E. H. Gelsight svelte: A human finger-shaped single-camera tactile robot finger with large sensing coverage and proprioceptive sensing. in *2023 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)* 8979–8984 (IEEE, 2023).

32. Sun, H., Kuchenbecker, K. J. & Martius, G. A soft thumb-sized vision-based sensor with accurate all-round force perception. *Nature Machine Intelligence* **4**, 135–145 (2022).
33. Ma, Y., Zhao, J. & Adelson, E. GelLink: A Compact Multi-phalanx Finger with Vision-based Tactile Sensing and Proprioception. (2024).
34. Li, S. *et al.* JamTac: A Tactile Jamming Gripper for Searching and Grasping in Low-Visibility Environments. *Soft Robotics* **10**, 988–1000 (2023).
35. Crooks, W., Vukasin, G., O’Sullivan, M., Messner, W. & Rogers, C. Fin ray® effect inspired soft robotic gripper: From the robosoft grand challenge toward optimization. *Frontiers in Robotics and AI* **3**, 70 (2016).
36. Liu, S. Q. & Adelson, E. H. Gelsight fin ray: Incorporating tactile sensing into a soft compliant robotic gripper. in *2022 IEEE 5th International Conference on Soft Robotics (RoboSoft)* 925–931 (IEEE, 2022).
37. Liu, S. Q., Ma, Y. & Adelson, E. H. Gelsight baby fin ray: A compact, compliant, flexible finger with high-resolution tactile sensing. in *2023 IEEE International Conference on Soft Robotics (RoboSoft)* 1–8 (IEEE, 2023).
38. Liang, S. *et al.* AllTact Fin Ray: A Compliant Robot Gripper with Omni-Directional Tactile Sensing. (2025).
39. Wang, S., She, Y., Romero, B. & Adelson, E. Gelsight wedge: Measuring high-resolution 3d contact geometry with a compact robot finger. in *2021 IEEE International Conference on Robotics and Automation (ICRA)* 6468–6475 (IEEE, 2021).
40. Liu, S. Q., Yañez, L. Z. & Adelson, E. H. GelSight EndoFlex: A Soft Endoskeleton Hand with Continuous High-Resolution Tactile Sensing. in *2023 IEEE International Conference on Soft Robotics (RoboSoft)* 1–6 (2023). doi:10.1109/RoboSoft55895.2023.10122053.
41. Liu, X., Han, X., Hong, W., Wan, F. & Song, C. Proprioceptive learning with soft polyhedral networks. *The International Journal of Robotics Research* 02783649241238765 (2024).
42. Chi, C. *et al.* Diffusion policy: Visuomotor policy learning via action diffusion. *The International Journal of Robotics Research* 02783649241273668 (2023).
43. Zitkovich, B. *et al.* Rt-2: Vision-language-action models transfer web knowledge to robotic control. in *Conference on Robot Learning* 2165–2183 (PMLR, 2023).
44. Chi, C. *et al.* Universal manipulation interface: In-the-wild robot teaching without in-the-wild robots. *arXiv preprint arXiv:2402.10329* (2024).
45. Wu, P., Shentu, Y., Yi, Z., Lin, X. & Abbeel, P. GELLO: A General, Low-Cost, and Intuitive Teleoperation Framework for Robot Manipulators. in *2024 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)* 12156–12163 (2024). doi:10.1109/IROS58592.2024.10801581.
46. Hansen, N., Müller, S. D. & Koumoutsakos, P. Reducing the time complexity of the derandomized evolution strategy with covariance matrix adaptation (CMA-ES). *Evolutionary computation* **11**, 1–18 (2003).

47. Nomura, M., Akimoto, Y. & Ono, I. CMA-ES with Learning Rate Adaptation. *ACM Transactions on Evolutionary Learning* (2024).

Acknowledgements

This work was supported in part by the National Natural Science Foundation of China through the Youth Program under Grant 509109-N72401; in part by the National Natural Science Foundation of China through the high-level Youth Talent Program under Grant 588020-X4240A/008; in part by the China Postdoctoral Science Foundation under Grant 2024M762814.

Ethics declarations

The authors declare no competing interests.