Chemical Engineering Journal 541 (2026) 177786

Contents lists available at ScienceDirect

Chemical Engineering Journal

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

Bionic *Tursiops Truncatus*-inspired dual-mode sensor for proximity sensing and high-resolution tactile perception

Hang Yu [a], Zhao Xu [a], Haowen Chen [a], Guangchao Wang [a], Qiang Sun [a], Haoran Fu [a], Huaping Wu [b], Aiping Liu [a,*]

[a] Key Laboratory of Optical Field Manipulation of Zhejiang Province, College of science, Zhejiang Sci-Tech University, Hangzhou, 310018, China
[b] Key Laboratory of Special Purpose Equipment and Advanced Processing Technology, Ministry of Education and Zhejiang Province, College of Mechanical Engineering, Zhejiang University of Technology, Hangzhou, 310023, China

ARTICLE INFO

Keywords:
Multimodal sensor
Proximity sensing
Tactile perception
Neural network model
Sensing and control system

ABSTRACT

Traditional vision-based tactile perception is one of the key enabling technologies for advancing machine intelligence, which captures and analyzes the optical deformation characteristics of the elastic body using a camera. However, the single-modal tactile sensor lacks integrated technical frameworks that can effectively combine high-density tactile sensing with proximity perception. This limitation restricts the development of comprehensive multimodal perception capabilities, especially during the pre-interaction phase, and thus fails to meet the requirements of high-precision, high-safety applications, such as surgical robotics and human-robot collaboration. To address this issue, this study presents a Visual – tactile - proximity (VTP) sensor, which is designed to overcome the limited proximity detection and insufficient tactile resolution. Through a gratifying composite structure, the VTP sensor integrates a visual-tactile sensing unit and a capacitive proximity sensing unit onto a single platform. The visual-tactile sensing unit can achieve a tactile sensing density of 16,640 units/cm^2, enabling precise three-dimensional surface reconstruction and fine feature identification upon contact. The capacitive proximity sensing unit adopts a multi-channel electrode configuration, providing an effective sensing range of up to 40 mm. It supports both the quantitative measurement of approaching object distances and qualitative material classification. Based on the VTP sensor, we developed a robotic arm sensing and control system. By integrating threshold-based decision algorithms with neural network models, this system can successfully perform active obstacle avoidance and intelligent object recognition. Nevertheless, the sensor has several drawbacks: limited adaptability to environmental temperature and humidity, slight mechanical fatigue after long-term use, susceptibility to nearby metallic interference, and low operational efficiency due to sequential data acquisition in the prototype. Experimental results indicate that the proposed sensor has significant potential for applications in human-robot collaboration and intelligent manufacturing, offering a satisfactory solution and critical technical support for enhancing multimodal perception in intelligent machines.

1. Introduction

As a core component that enables intelligent robots to achieve environmental interaction and precision manipulation, tactile sensors not only endow robots with perception capabilities surpassing human abilities but also serve as the key technological support for the upgrade of traditional mechanical equipment toward safety, precision, and intelligence [1–5]. Currently, tactile sensors based on different sensing mechanisms have been comprehensively studied [6]. Among them, thanks to hardware technological innovation and rapid development of image processing algorithms, vision-based tactile sensors have become the economically optimal solution for achieving high-resolution tactile perception [7–10]. However, most existing visual-tactile sensors primarily rely on the contact between objects and the sensor surface, which causes the deformation of the elastic material, thus facilitating the acquisition of tactile information [8]. Additionally, these sensors are mainly based on the principle of optical imaging, which involves capturing the deformation image of the elastic body via a camera and analyzing the physical information [11], though the ability to recognize the material properties of the object is relatively limited. In the context

* Corresponding author.
E-mail address: liuaiping1979@gmail.com (A. Liu).

https://doi.org/10.1016/j.cej.2026.177786
Received 24 January 2026; Received in revised form 12 May 2026; Accepted 26 May 2026
Available online 27 May 2026

of the in-depth progress toward intelligence and multifunctionality in fields such as medical care, human-robot interaction, and precision manufacturing, traditional single-modal tactile sensors are restricted to contact signal perception and cannot predict potential dangers during the pre-interaction phase, thereby failing to meet the requirements of high-precision, high-safety applications. Therefore, developing dual-mode tactile sensors that integrate high-resolution tactile perception with proximity sensing has become a critical technical challenge.

The vision-based tactile sensors equipped with proximity perception function overcome the limitation of solely relying on contact perception. They are capable of perceiving the presence, distance, and motion state of objects when the objects approach but do not actually make contact, thereby expanding the perception range and enabling robots to respond to surrounding objects in advance [12–14]. Owing to the implementation of multimodal sensor fusion technology, multiple sensing methods complement and verify one another, which enhances the anti-interference ability and perception stability in complex environments [15–19]. Through preprocessing multimodal signals and employing feature extraction algorithms for signal fusion analysis and processing, more efficient and accurate identification of the surface material properties of objects can be achieved [20]. Currently, the academic community has carried out a variety of studies on integrating visual-tactile perception with proximity sensing functionality. For example, Dong et al. [12] put forward an integrated sensor based on a mechanically driven active mode-switching mechanism. Nevertheless, this solution has drawbacks such as complex and fragile mechanical structures, high manufacturing costs, long mode-switching delays, and substantial computational overhead, which makes it hard to meet long-term stable application requirements in practical scenarios. Yin et al. [13] achieved simultaneous acquisition of proximity depth data and tactile depth data by integrating infrared time-of-flight (TOF) cameras with stereo cameras. However, this system has problems including structural complexity, large size, high integration difficulty, high cost, and inadequate algorithm stability. Mu et al. [14] constructed a proximity-contact dual-modal sensor based on triboelectric sensor arrays and finger-shaped visual-tactile sensors. The sensors were capable of sensing dynamically spatial movement of external objects without contact and could generate electrical signals containing material property information through friction, thereby achieving the synergistic effect of different sensing technologies. Yet, the fabrication processes of three-dimensional structured visual-tactile sensor is intricate, and the three-dimensional reconstruction imaging from tactile raw images presents significant challenges.

To enable robots to acquire relevant information during object

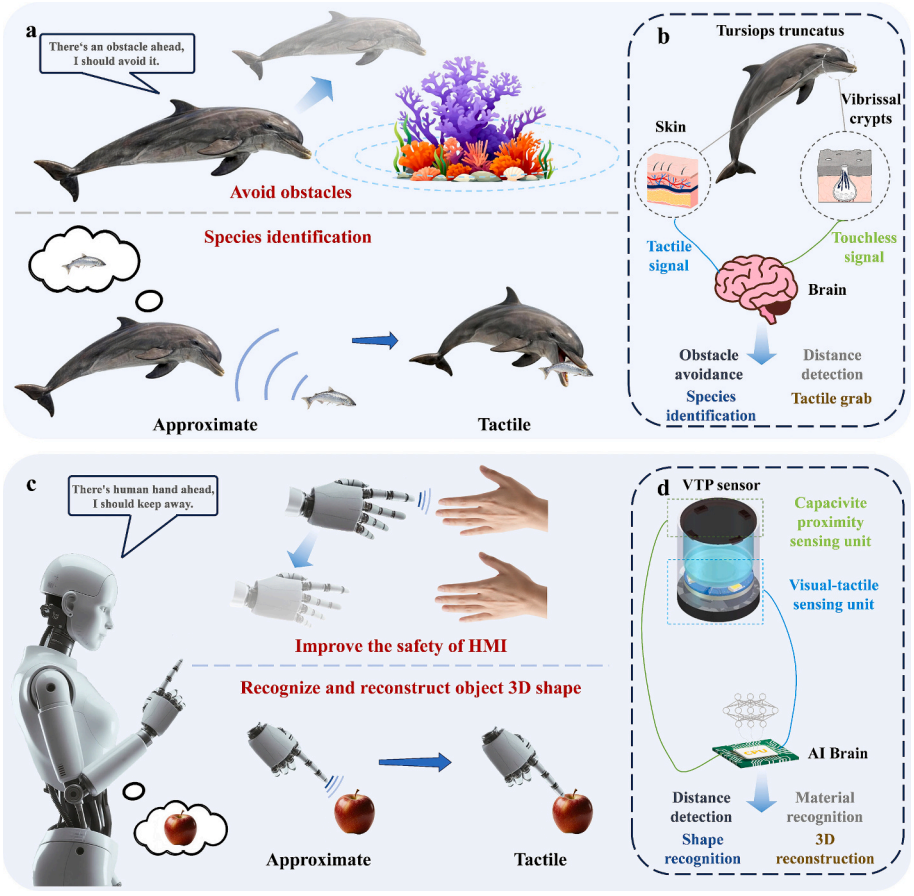

Fig. 1. Conceptual diagram of a multimodal perception system inspired by the integration of electroreceptive and tactile senses of *Tursiops truncatus*. (a) In benthic foraging environments with turbid water, *Tursiops truncatus* can achieve close-range obstacle avoidance (less than 10 cm) and hidden prey recognition (3–7 cm) via electroreceptors. Subsequently, they complete prey capture via tactile sensing; (b) The multimodal sensory mechanism of *Tursiops truncatus*: the vibrissal crypts serve as electroreceptors to receive non-contact signals, and the high-density nerve endings distributed in the skin act as mechanoreceptors to detect tactile stimuli. Both types of signals are then transmitted to the brain to synergistically fulfill functions such as distance perception and obstacle avoidance, prey identification, and grasping behaviors; (c) Working principle of biomimetic fingertip sensor: capacitive proximity signals enable obstacle avoidance and material identification, while visual-tactile signals accomplish tactile perception and grasping operations; (d) Composition of the biomimetic multimodal sensor (VTP sensor): includes a capacitive proximity sensing unit and a visual-tactile sensing unit, with the former responsible for non-contact signal acquisition and the latter for tactile signal capture. Both types of signals are transmitted to the central processing unit (CPU) and combined with neural network algorithms to achieve functions including distance perception, material identification, shape recognition, and 3D reconstruction.

approach, pre-plan operation strategies, pre-perceive the position, posture, and distance of objects during grasping, adjust grasping paths and forces, and ultimately improve grasping success rates and operational efficiency, inspiration is drawn from the electroreceptive physiological characteristics of *Tursiops truncatus*. These marine mammals detect biological and non-biological obstacles within a centimeter- to decimeter-scale range ahead through electroreceptors in the highly innervated vibrissal crypts on their rostrum. Endogenous electric fields from biological targets are directly detectable, while non-biological targets, although they do not generate electric fields themselves, induce perturbations in the ambient electric field to form "electric field distortion" signals—these signals can be captured by the dolphins' highly sensitive electroreceptive system for indirect identification. Thanks to this mechanism, *Tursiops truncatus* can not only achieve close-range obstacle avoidance and environmental perception but also accomplish precise predation by detecting weak bioelectric fields emitted by the gills of prey hidden in sediments (Fig. 1a-b) [21,22]. Based on the above bionic mechanism, this study proposes an integrated-manufactured VTP sensor. Through a composite structural design, the VTP sensor achieves high level of integration between capacitive proximity sensors and visual-tactile sensors, effectively overcoming defects of existing technologies including complex and fragile mechanical structures, high cost, poor integration, the inability to sense static approach signals, and difficulties in three-dimensional reconstruction. While guaranteeing high-density tactile perception imaging and high-precision proximity sensing performance, the sensor is characterized by a compact structural design and can be conveniently integrated into robot fingertips or robotic gripper jaws. The robotic arm sensing and control system developed based on the VTP sensor can fully utilize its proximity-contact dual-mode sensing advantages to achieve active obstacle avoidance and multi-mode object recognition functions of gripper, offering high-performance sensing solutions for fields like human-robot collaboration and intelligent manufacturing (Fig. 1c-d).

2. Experimental methods

2.1. Fabrication of dual-mode tactile sensor

The fabrication process of the VTP sensor was as follows: First, the sensor housing and base were designed via 3D modeling and were fabricated using 3D printing technology. The camera module and LED ring were fixed to the base. 1-mm-thick acrylic sheets were processed via infrared laser cutting and were installed inside the housing. Liquid PDMS (Sylgard 184, Dow Corning, USA, with a ratio of 10: 1) was poured into the housing, vacuum degassing was performed for 5 min, and it was then heat-cured at 60 °C for 3 h to achieve seamless bonding between the acrylic sheets and the PDMS support structure. Ecoflex (Sylgard 184, Dow Corning, USA, with a ratio of 1:1) was mixed with multi-walled carbon nanotubes (MWCNTs, 95% purity) at a mass ratio of 95%:5%, vacuum degassing was conducted, the mixture was poured into molds and cured to fabricate the lower electrode of the capacitive proximity sensing unit, and the electrode was placed on the PDMS layer. Semi-transparent liquid silicone with a Shore hardness of 5 A (DRSGJ02, POSILICONE Company, with a ratio of 1:1) was poured into the housing, vacuum degassing was carried out, and it was cured at room temperature. Finally, the Ecoflex/CNTs mixed solution ($W_{Ecoflex}$:W_{CNTs} = 95%:5%) was poured, vacuum degased, and cured at room temperature, and the assembly of the housing and base was completed. The detailed fabrication process was shown in Fig. S1.

2.2. Characterization and measurement

A scanning electron microscope (SEM, CARL ZEISS SMT PTE Ltd.) was employed to analyze the microstructure of CNTs/Ecoflex at an ultra-high voltage of 3 kV. A mechanical testing machine (MX-0350, Jiangsu Moxin Industrial Systems Co., Ltd.) was used to simulate the processes of uniform object approach, retreat, and normal pressing. A digital bridge (TH2830, Changzhou Tonghui Electronics Co., Ltd.) was utilized to collect capacitive sensing signals, with the data display and storage were accomplished via LabVIEW software. An OV2640 camera module (with a resolution of 480 × 320 resolution and a frame rate of 30 frames/s) was used to collect tactile images, which were transmitted to the computer via WIFI communication with image data processing and storage completed through Python programs. In practical applications, the data acquisition process of the entire system encompassed optical imaging and multi-channel capacitance acquisition. Accordingly, two circuit boards were designed to facilitate clear imaging and data acquisition in the system, as depicted in Fig. S2 and Fig. S3.

2.3. Finite element analysis

To precisely clarify the influence patterns of electrode structure and dimensions on the electric field distribution of the capacitive sensing unit, finite element analysis (FEA) was conducted using COMSOL Multiphysics software. Based on the software's built-in electrostatic field module, a three-dimensional simulation model of the capacitive sensing unit was constructed. This model had a geometric structure precisely matching the actual sensor, including an upper plate, an insulating dielectric layer, and four lower plates. During simulation, the electrical conductivity of electrode materials and relative permittivity of the insulating dielectric layer were accurately defined according to actual material parameters to ensure physical authenticity of the simulation model. To systematically investigate the influence of electrode structure and dimensions, various electrode structure models including disk-shaped, ring-shaped, and maze-shaped ones were established. Moreover, disk-shaped electrode models with different radius specifications were also created. A comprehensive simulation analysis of potential distribution and electric field strength distribution was conducted for each type of model. The simulation boundary conditions were set as follows: A 1-V DC voltage was applied to the upper plate as the common excitation electrode. The four lower plates were connected in parallel and grounded to 0 V ground. The simulation calculation domain was extended to the air domain to completely characterize the electric field propagation characteristics around the sensor. Finally, through numerical calculation, electric field distribution cloud maps and quantitative data for the capacitive sensing unit in the air domain under different structures and dimensional parameters were obtained. These results could offer theoretical support for the optimization design of electrode structure.

2.4. Statistical analysis

To achieve precise object classification based on VTP sensor's sensing signals, this study constructed a multi-mode recognition system based on neural networks. First, a detailed process of dataset construction and partitioning was implemented. For the material classification task based on capacitive time-series signals, a total of 2400 samples were collected, covering 8 material categories with 300 samples per category. Each capacitive sample is a 4 × 398 single-channel two-dimensional vector. For the tactile image classification task of 26 English letters, 1300 image samples were acquired, with 50 samples for each letter category. After preprocessing, each image sample is a 16 × 16 single-channel two-dimensional vector. All samples were divided into training, validation, and test sets via stratified random sampling with a ratio of 7:1:2, ensuring consistent class distribution across subsets for a more robust model evaluation.

For the image data collected by the visual-tactile sensing unit, data preprocessing was initially performed using the OpenCV open-source vision library. The original RGB images with a resolution of 480 × 320 were rotated and cropped to a resolution of 200 × 200 for interface display, then converted into single-channel grayscale images by averaging the 8-bit pixel values of the three RGB channels, and finally

downsampled to 16 × 16-pixel single-channel grayscale images. To avoid information leakage, strict data partitioning was implemented prior to any preprocessing and augmentation operations. Basic preprocessing steps (rotation, cropping, grayscale conversion, and downsampling) were applied equally to the training, validation, and test sets. Data augmentation (including rotation, position offset, and brightness adjustment) was exclusively applied to the training set, while the validation and test sets were not subjected to any augmentation. All normalization and feature extraction procedures were conducted independently on each subset, thereby eliminating the risk of information leakage.

Subsequently, a 7-layer convolutional neural network (CNN) model was developed based on the PyTorch deep-learning framework. The model takes 16 × 16 single - channel grayscale images as input, and the output layer is configured with 26 neurons, which correspond to the classification task of 26 English letters. During the model training process, the cross-entropy loss function was employed to measure classification error. The Adam optimizer was selected for parameter updates, with the initial learning rate set at 0.001, and total training epochs set to 25 (The training process was shown in Fig. S4). To enhance training efficiency, the entire training process was executed on an NVIDIA GeForce MX150 GPU, reducing the model's convergence time via hardware acceleration.

For the capacitance time-series data collected by the capacitive proximity sensing unit, baseline calibration was first conducted to eliminate signal drift: 10 groups of consecutive capacitance data were averaged as the baseline value, and all subsequent sampled values were subtracted from this baseline to obtain valid capacitance variations. Considering the noise robustness of neural networks and real-time sensing requirements, no additional filtering was applied. The 4 × 398 two-dimensional capacitive array was flattened into a one-dimensional vector and normalized, forming a 1592-dimensional time-series feature vector as the model input for the model.

A multi-layer perceptron (MLP) model with 3 hidden layers was constructed, with an output layer of 8 neurons to match the 8-material classification task. During model training, the cross - entropy loss function and the Adam optimizer were also employed. To adapt to the feature distribution characteristics of capacitance data, the initial learning rate was adjusted to 0.0001, and the total number of training epochs was set to 250. This adjustment could ensure that the model fully learned the capacitance change pattern features corresponding to different materials (The training process was shown in Fig. S5).

2.5. Sensing and control system

To integrate the sensor into a robotic arm system for executing real-time operational tasks such as active obstacle avoidance and intelligent sorting, the sensing and control system for the robotic arm was designed. First, the structural design and fabrication of the robotic gripper integrated with VTP sensor was completed, ensuring reliable assembly with the 6-DOF robotic arm through flange connectors and M3 specification fastening screws. In the design of the sensing signal acquisition and transmission link, the FDC2214 capacitance-to-digital conversion chip sampled the VTP sensor's capacitance signals and transmited them to the ESP32 via I2C communication protocol. Visual-tactile image data was collected by the OV2640 camera module and sent to ESP32 via SCCB bus through serial transmission. Subsequently, data uploading was achieved through dual communication channels: The capacitance data was transmitted serially, and the image data was transmitted via WIFI communication, which were ultimately synchronized to the computer to complete real-time acquisition, storage, and preprocessing of sensing data. The motion control of the robotic arm employed STM32 as the main control unit. The control program was written and debugged based on the Keil software, enabling real-time motion control of the robotic arm joints through commands. To achieve visualized monitoring and convenient operation of the system, a human-machine interaction interface (GUI) was designed based on the QT development platform for displaying the real-time sensing data of VTP sensor and supporting real-time control of the robotic arm.

To ensure reliable and stable operation in robotic applications, a comprehensive anti-interference design and hardware grounding optimization were implemented for the sensing and control system. Except for the servo and screws and nuts, which are located far from the sensing surface and are made of metal materials, all other structural parts of the gripper are fabricated using 3D-printed resin or nylon, and the gripping contact surface is made of TPU. These non-metallic materials effectively reduce the intrinsic electromagnetic interference from the gripper body. Meanwhile, the motion trajectory of the robotic arm is programmed to maintain a safe distance between the sensor and other metal components during operation. Experimental verification shows that the capacitance fluctuation is controlled within 0.01 pF during the full-range opening and closing of the gripper. The robustness of the neural network algorithm will not be affected by such weak noise of 0.01 pF, and therefore, it will not impact the final accuracy of material recognition and obstacle avoidance.

The system adopts an optimal single-point grounding configuration without ground-loop interference. The complete grounding topology is as follows: the capacitive sensor is connected to the FDC2214 acquisition board, and the GND of the FDC2214 capacitance acquisition board is connected to the GND of the STM32 control board; the STM32 is powered by a two-hole power adapter without direct connection to the laboratory earth; the entire system is connected to the laboratory earth solely via the path of STM32 GND-CH340 GND-computer GND. To further validate the engineering adaptability in industrial scenarios, two additional typical grounding configurations were simulated. Under the dual-point grounding (ground-loop) condition, the STM32 GND is connected to the laboratory earth through a Dupont wire. In the floating grounding condition, the CH340 USB cable is unplugged, resulting in the whole system electrically floating. These two conditions will introduce noise and cause drift in the baseline of the sensor, and thus are not adopted.

3. Results and discussion

3.1. Structural design and operating principle of VTP sensor

The core design objective of VTP sensor is to achieve integrated fusion of the capacitive proximity sensing unit and visual-tactile sensing unit via a composite structural design, with the specific structure presented in Fig. 2a. The capacitive proximity sensing unit adopts a four-channel capacitive sensor array architecture, which consists of an upper common electrode, a semi-transparent elastomer dielectric layer, and four lower electrodes. The fabrication process of these electrodes is illustrated in Fig. 2b. The visual-tactile sensing unit is composed of a miniature camera module, a light-emitting diode (LED) ring, a semi-transparent elastomer, and auxiliary support components (Fig. S6). Among these components, the semi-transparent elastomer acts as a crucial functional layer fulfilling a dual role: firstly, it reflects the light emitted by the LED ring in the visual-tactile sensing unit, providing an optical foundation for tactile imaging [23,24]; secondly, it functions as the dielectric layer of the capacitive proximity sensing unit, ensuring the performance of capacitive sensing. CNTs exhibit excellent mechanical and electrical properties. When compounded with Ecoflex, they are used to fabricate the elastomer electrodes of the VTP sensor [25]. The SEM images of CNTs/Ecoflex (as shown in Fig. 2d and e) reveal that these elastomers possess wrinkled surfaces, demonstrating their flexibility and laying a solid foundation for the excellent imaging performance of the visual-tactile sensing unit. Additionally, the upper electrode of the capacitive proximity sensing unit also serves as the outermost structure of the visual-tactile sensing unit. It can not only absorb the internal light that penetrates through the semi-transparent elastomer but also effectively prevent ambient stray light from interfering with the internal

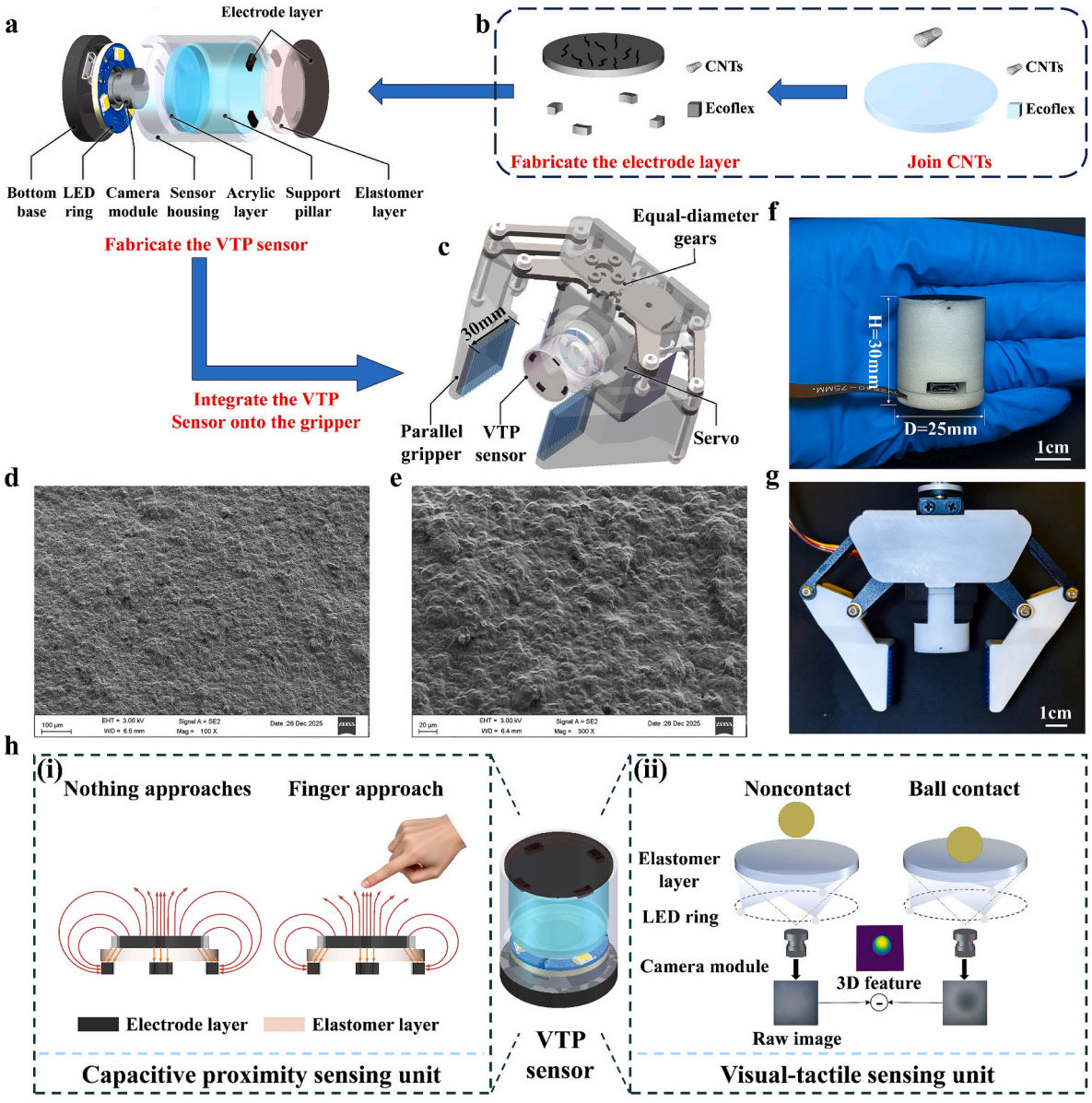

Fig. 2. Structural design and operating principle of VTP sensor. (a) Integrated structural design of VTP sensor: the integrated fusion of capacitive proximity sensor and visual-tactile sensor is achieved through functional layer reuse strategy; (b) Fabrication process of CNTs/Ecoflex electrode layer; (c) Structural design of a robotic gripper integrated with VTP sensor, with effective gripping range of 7–37 mm in front of VTP sensor; (d and e) SEM images at different magnifications of CNTs/Ecoflex; (f) Optical photograph of VTP sensor, showing compact and small structure suitable for integration applications in robotic grippers; (g) An optical photograph of robotic gripper integrated with VTP sensor; (h) Schematic diagram of VTP sensor operating principle: (i) Electric field coupling mechanism of capacitive proximity sensing unit; (ii) Optical imaging and reconstruction mechanism of visual-tactile sensing unit.

optical path, realizing efficient reuse of the functional layer. Fig. 2f shows the fabricated VTP sensor with a height of merely 30 mm and a circular cross-section diameter of 25 mm, which is easy to be integrated into a servo-driven mechanical gripper (Fig. 2c). The effective working range of its parallel gripping surfaces is 7–37 mm in front of the VTP sensor (Fig. S7), which is in excellent agreement with the proximity sensing range of the VTP sensor (Fig. 3f). To avoid interfering with the VTP sensor itself, all mechanical components of the gripper are fabricated from resin materials using 3D printing technology (Fig. 2g). The gripper structure is stable and reliable, and it is expected to stably perform complex operational tasks such as human-robot interaction and intelligent sorting after integration with the robotic arm.

The VTP sensor achieves a comprehensive perception of both proximity and contact states through coordinated operation of the capacitive proximity sensing unit and visual-tactile sensing unit. The total capacitance of the capacitive proximity sensing unit comprises two components: the mutual capacitance from the upper electrode passes through the dielectric layer to the lower electrode, and the fringe capacitance of diverging electric field lines from the upper electrode does not pass through the dielectric layer (Fig. 2h(i)). In the proximity state, the sensor can detect the approach behavior of external objects by leveraging the fringe field effect of the capacitive proximity sensing unit [26]. When an external object approaches the sensor, the fringe field lines are intercepted, causing a decrease in fringe capacitance and consequently a change in total capacitance value. By detecting the pattern of capacitance change, proximity sensing of objects can be achieved [27]. The electric field distribution characteristics of the capacitive proximity sensing unit can be visually presented by conducting a finite-element simulation analysis of its electrostatic field distribution using COMSOL software (Fig. 3e). In the contact state, the

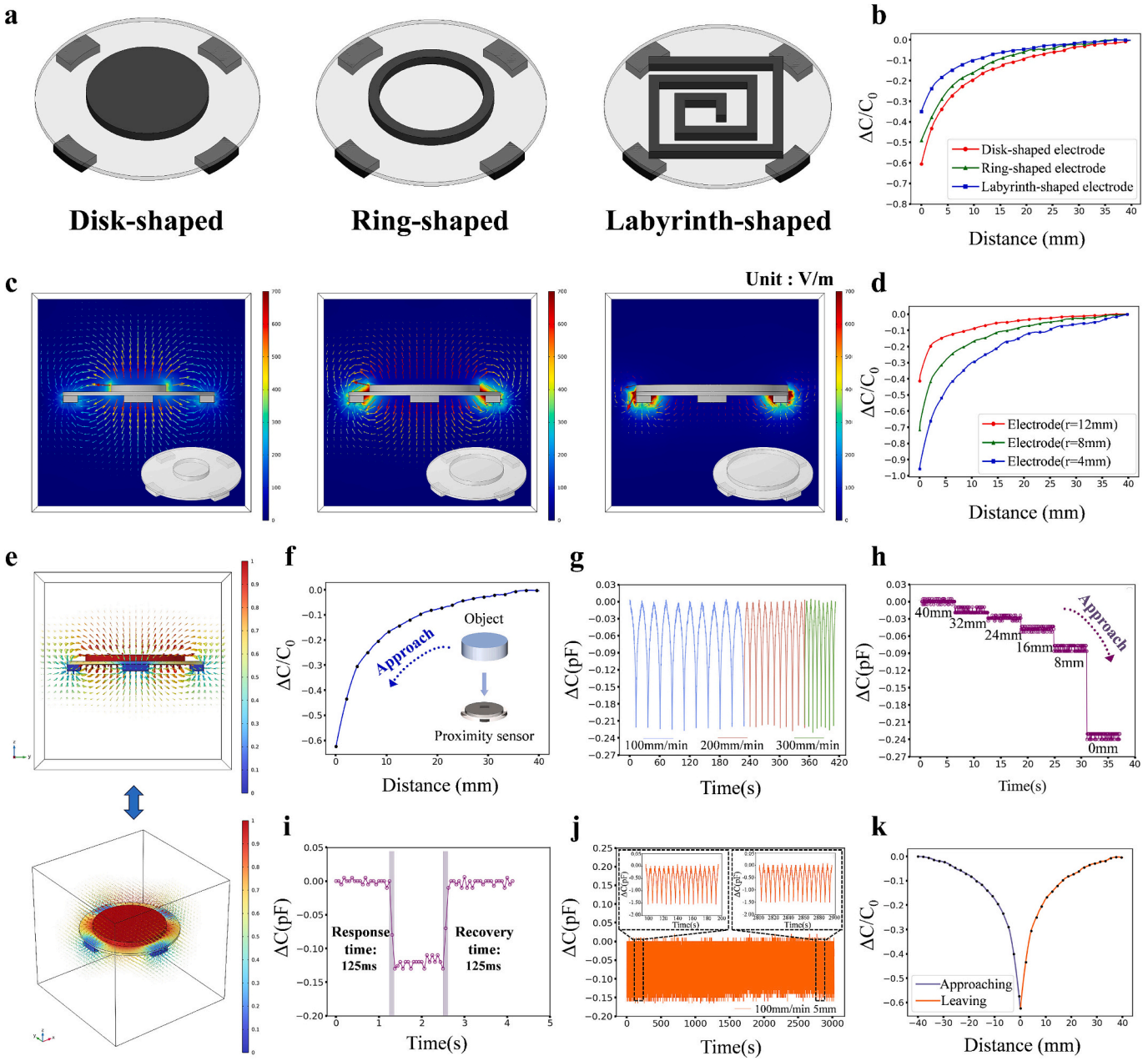

Fig. 3. Optimization of electrode structure and size and testing of sensing performance for the proximity sensing unit of VTP sensor. (a) Design schemes of three electrode structures (disk-shaped, maze-shaped, ring-shaped) along with the corresponding simulation results of electrostatic field distribution; (b) Curves depicting the relationship between capacitance change and the object's approach distance for different electrode structures; (c) 2D Cross-Section: Simulation results of electrostatic field distribution for disk-shaped electrodes with different radii (4 mm, 8 mm, 12 mm); (d) Curves showing the relationship between capacitance change and the object's approach distance for disk-shaped electrodes with different radii; (e) 3D Distribution Map: Simulated electrostatic field distribution for the disk-shaped electrode with 8 mm radius; (f) Relative capacitance change curves of capacitive proximity sensing unit of VTP sensor; (g) Capacitance response characteristics of capacitive proximity sensing unit of VTP sensor at different approach velocities; (h) Signal stability testing of the capacitive proximity sensing unit of VTP sensor at different approach distances; (i) Test curves for response time and recovery time of the capacitive proximity sensing unit of VTP sensor; (j) Repeatability curves of the capacitive proximity sensing unit of VTP sensor over 500 cycles; (k) Hysteresis test curves of the capacitive proximity sensing unit of VTP sensor.

visual-tactile sensing unit achieves tactile perception through the deformation of the elastomer [23,24]. When an object comes into contact with the elastic sensing body, the resulting deformation modifies the propagation path of light emitted by the LED ring, causing the built-in miniature camera to capture images that are significantly different from those in the non-contact state. Tactile signal extraction can be accomplished through image feature analysis (Fig. 2h(ii)).

3.2. Sensing performance of VTP sensor

Since proximity detection is accomplished via capacitive coupling between the fringe field generated above the capacitive proximity sensing unit and objects, proximity sensing performance is closely related to electrode structure and dimensions [16]. Here, the electrode structure and dimensions of the capacitive proximity sensing unit were initially optimized through a combination of finite-element simulation and experimental testing. The COMSOL software was employed to perform a static simulation of the electric field distribution for three

electrode structures: disk-shaped, ring-shaped, and square maze-shaped (Fig. 3a). The results indicate that the range and gradient of the fringe field generated by the disk-shaped electrode structure are significantly superior to those of the other two structures. To verify the simulation results, sensor samples with the three electrode structures were fabricated and underwent comparative experiments. When objects approach uniformly from 40 mm, the relative capacitance change of the disk-shaped electrode structure exceeds 60%, which is far higher than the capacitance change response of the maze-shaped and ring-shaped structures (Fig. 3b), in line with the simulation results. Regarding the optimization of electrode size, electrostatic field distribution simulations were performed for disk-shaped electrodes with radii of 4 mm, 8 mm, and 12 mm (Fig. 3c). It is found that the 12-mm-radius electrode only generates weak electric fields at the sides, while the 4-mm and 8-mm radius electrodes form fringe fields of different intensities and ranges above the sensors. Experimental testing demonstrates that the electrode with a 4 mm radius exhibits the largest relative capacitance change (>90%, Fig. 3d). However, owing to its small initial capacitance value and consequently a small capacitance variation, it leads to insufficient discrimination of the sensing signal in practical applications (Fig. S8). By integrating the simulation and experimental results, a disk-shaped electrode with 8 mm radius is ultimately selected as the optimal structural solution for the capacitive proximity sensing unit, and its detailed electrostatic field distribution simulation (shown in Fig. 3e) reveals that: between the common electrode and the four underlying small electrodes of this structure, a high-field-strength mutual capacitance region is formed, while a significant fringe electric field extends around the periphery of the common electrode; this fringe field can be disturbed by approaching external objects and induce capacitance changes, thus enabling this structure to function as a proximity sensor for object proximity detection.

Systematic sensing performance testing was conducted on the optimized proximity sensing unit of VTP sensor, with the results shown in Figs. 3f-k. The unit's effective sensing range reaches 40 mm, with relative capacitance change gradually increasing as the object approach distance decreases (Fig. 3f). When the distance approaches 0 mm, the relative capacitance change reaches 62.4%. As the objects approach to the sensor from a distance of 30 mm at three velocities of 100 mm/min, 200 mm/min, and 300 mm/min through the approach-retreat process, the capacitance response presents cyclically changing for moving objects at different velocities (Fig. 3g). At a given spacing distance, the sensing signal remains stable (Fig. 3h), thus verifying the reliable perception capability of the sensor. The long-term static capacitance stability was evaluated by keeping the sensor in a stationary state for 3 h (Fig. S9). The sensor exhibits outstanding static signal stability with a noise level of less than 0.01 pF and no obvious capacitance drift, which guarantees the reliability of long-term proximity signal acquisition. Response time testing (Fig. 3i) indicates that both response time and recovery time are 125 ms, which meets the requirements for real-time detection and high-frequency signal transmission. The results of 500 approach-retreat cycle tests (Fig. 3j) and hysteresis testing (Fig. 3k) demonstrate that the sensor maintains stable performance during long-term use, with basically consistent capacitance change during approach and retreat processes, indicating excellent hysteresis performance. Furthermore, a 5000-cycle pressing fatigue test was implemented to assess the elastomer fatigue and long-term tactile sensing robustness (Fig. S10). The sensor was continuously pressed to a depth of 1.5 mm per cycle, and its sensing performance was calibrated prior to the test and after every 1000 cycles. A slight performance degradation was observed as the number of cycles increased, which is ascribed to the minor mechanical fatigue of the soft elastomer layer. The environmental adaptability of the VTP sensor was systematically characterized under varied temperature and humidity in a self-built enclosed temperature-humidity control chamber (Fig. S11). At a fixed relative humidity of 50 ± 2.5% RH, the proximity sensing performance continuously deteriorates with temperature increasing from 25 ± 0.5 °C to 60 ± 2.5 °C, showing a maximum performance attenuation of approximately 21%. At a fixed temperature of 25 ± 0.5 °C, the sensing performance gradually weakens as relative humidity rises from 50 ± 2.5% to 80 ± 2.5% RH, with a maximum attenuation of about 8%. Such performance degradation is ascribed to the distorted fringe electric field caused by temperature-induced variations in dielectric properties, electrode conductivity and structural thermal expansion, as well as humidity-driven water adsorption, conductive network disturbance and substrate swelling. These results demonstrate the obvious environmental sensitivity of the sensor and offer essential support for the design of environmental compensation strategies and its practical implementation in complex robotic application scenarios.

Furthermore, the anti-interference robustness of the VTP sensor was quantitatively evaluated under practical working conditions. Regarding the mutual interference between the capacitive proximity unit and the visual-tactile imaging unit, turning the ESP32CAM microcontroller on or off has no impact on the capacitive sensing signal. When the built-in LED ring is powered on, it reduces the baseline capacitance of all channels by approximately 0.02 pF, yet the relative capacitance change rate remains fully unchanged (Fig. S12); the baseline returns to the original level immediately after the LED ring is powered off. Since the relative capacitance variation remains consistent, such a synchronous baseline shift does not affect the sensing performance or system decision-making. Meanwhile, the visual-tactile unit, which relies on optical imaging, shows high immunity to electromagnetic interference; no obvious change in the average grayscale of tactile images is observed before and after the capacitive sensing channel is powered on or off, confirming that the capacitive unit does not interfere with the visual-tactile unit. In actual operation, all electronic components are powered simultaneously during the entire working period, which completely eliminates intermittent mutual interference between the two modules. To address the interference from nearby metal objects, a 5 cm × 5 cm × 5 cm metal block placed parallel to the sensing surface was tested (Fig. S13). When the metal block is in close contact with the sensor (distance = 0 mm), the extreme value of the sensor's capacitance change decreases from 0.62 to 0.58, representing a reduction of approximately 6% in the capacitance variation amplitude.

The capacitive proximity sensing unit operates based on the electric field coupling effect. As a high-conductivity material, when the metal block approaches the sensor, it generates electrostatic induction. This electrostatic induction causes the metal block to absorb and shield part of the fringe electric field lines generated by the sensor electrodes. As a result, the capacitance component is reduced, leading to an overall downward offset of the baseline capacitance. Simultaneously, the local distortion of the electric field distribution leads to a slight variation in the relative capacitance change, which might affect the stability of subsequent algorithms. Nevertheless, this interference is highly distance-dependent and nearly disappears completely when the metal block is moved to a distance of 20 mm from the sensor end, thus ensuring safe operation within the designed working range.

Moreover, the contact sensing performance of the visual-tactile sensing unit of VTP sensor was systematically characterized by employing a tensile testing machine equipped with a force sensor (Fig. 4a). Sensing density, regarded as a core performance indicator, is defined as the quantity of effective sensing units per square centimeter (units/cm^2). By vertically pressing the sensor surface with a square probe of a known area and utilizing contour detection algorithms to count the number of pixels in the 10 mm × 10 mm actual contact area (approximately 129 × 129 pixels), the sensing density of the visual-tactile sensing unit was calculated to be approximately 16,640 units/cm^2 (Fig. S14), which far exceeds the 2400 units/cm^2 of human fingers, thereby providing a hardware foundation for high-density tactile perception and 3D reconstruction. Fig. 4b presents the force characteristics of the sensor at different pressing depths and the corresponding tactile raw images. Analysis results show that when the pressing depth is less than 1 mm, the discrimination of tactile image feature is low,

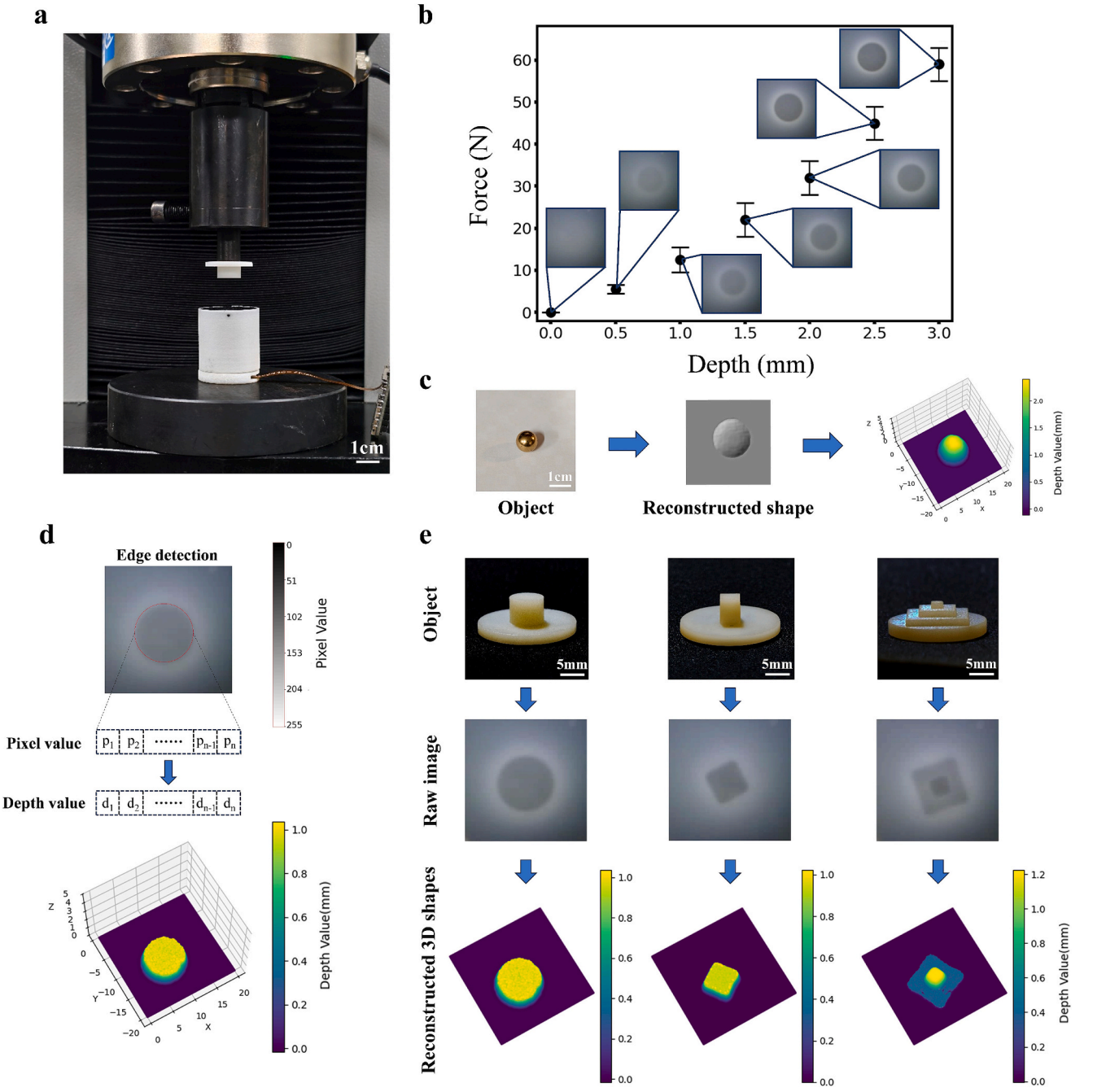

Fig. 4. Tactile sensing and 3D reconstruction capability testing of visual-tactile sensing unit of VTP sensor. (a) Experimental setup for the fixed-depth pressing test: a tensile testing machine equipped with a force sensor is used; (b) Relationship curves between pressing depth and force, along with tactile raw images at different depths; (c) Schematic diagram of the 3D reconstruction process for spherical objects; (d) Schematic diagram of the principle of 3D reconstruction of the tactile surface; (e) 3D reconstruction results for objects with different geometric shapes (cylinder, rectangular prism, square step block): optical images, tactile images, and reconstructed 3D contact geometric shapes are shown from top to bottom.

making it difficult to extract effective geometric information. When the pressing depth exceeds 2 mm, the required pressing force increases significantly, which can easily cause sensor damage. After comprehensively considering the imaging quality and operational safety, the optimal pressing depth range is determined to be 1–2 mm. Within this range, the sensor can obtain tactile raw images with clear features under small pressing forces, offering high-quality data support for subsequent 3D reconstruction. Fig. 4c demonstrates the 3D reconstruction process when spherical objects are randomly pressed on the visual-tactile sensor surface. Notably, based on contact mechanics theory and previous experimental data, the spherical indenter can achieve the target depth under a relatively small load. This characteristic enables spherical objects to capture more subtle mechanical features in precision sensing applications, thereby enhancing the detail performance of 3D reconstruction [28].

In contrast, flat-ended objects exhibit a constant contact area, with stress consistently concentrated in fixed regions, and thus require a larger load to generate equivalent deformation. Therefore, during the reconstruction process, it is necessary to focus on the extraction of surface detail features. Fig. 4d shows the principle diagram of 3D reconstruction of tactile surface. The 3D geometric shape reconstruction process for the visual-tactile sensing unit is as follows. First, tactile

grayscale images during object contact and baseline grayscale images in the non-contact state are separately collected, and difference images characterizing elastomer deformation are obtained through pixel-level difference calculation. Subsequently, by using pre-calibrated grayscale-depth lookup tables, the pixel grayscale values of difference images are mapped to depth coordinates in 3D space. Eventually, point cloud data for the contact region is generated, and the depth information reconstruction of the contact surface is achieved [23,24]. The detailed workflow is illustrated in Fig. S15 and Fig. S16. To quantitatively evaluate the 3D reconstruction accuracy of the visual-tactile sensing unit, a tensile testing machine was used to press cylindrical samples of different sizes to preset depths while keeping them stable. A total of 12 tactile image samples were collected. Contour detection algorithms were used to extract the effective pixels within the pressed region of each image, and reconstruction depth values corresponding to each pixel were calculated based on calibration lookup tables. By comparing these values with the true depth values set by the tensile testing machine, the mean absolute error (MAE) for reconstruction depth within the effective contact region was 0.0573 mm with a standard deviation (Std) of 0.0415 mm (The detailed calculation process is shown in Fig. S17). This finding confirms the high-precision 3D reconstruction capability of the visual-tactile sensing unit. To further validate the sensor's reconstruction adaptability to complex geometric shapes, three typical geometric shape samples were selected: a cylinder, a rectangular prism, and a square step

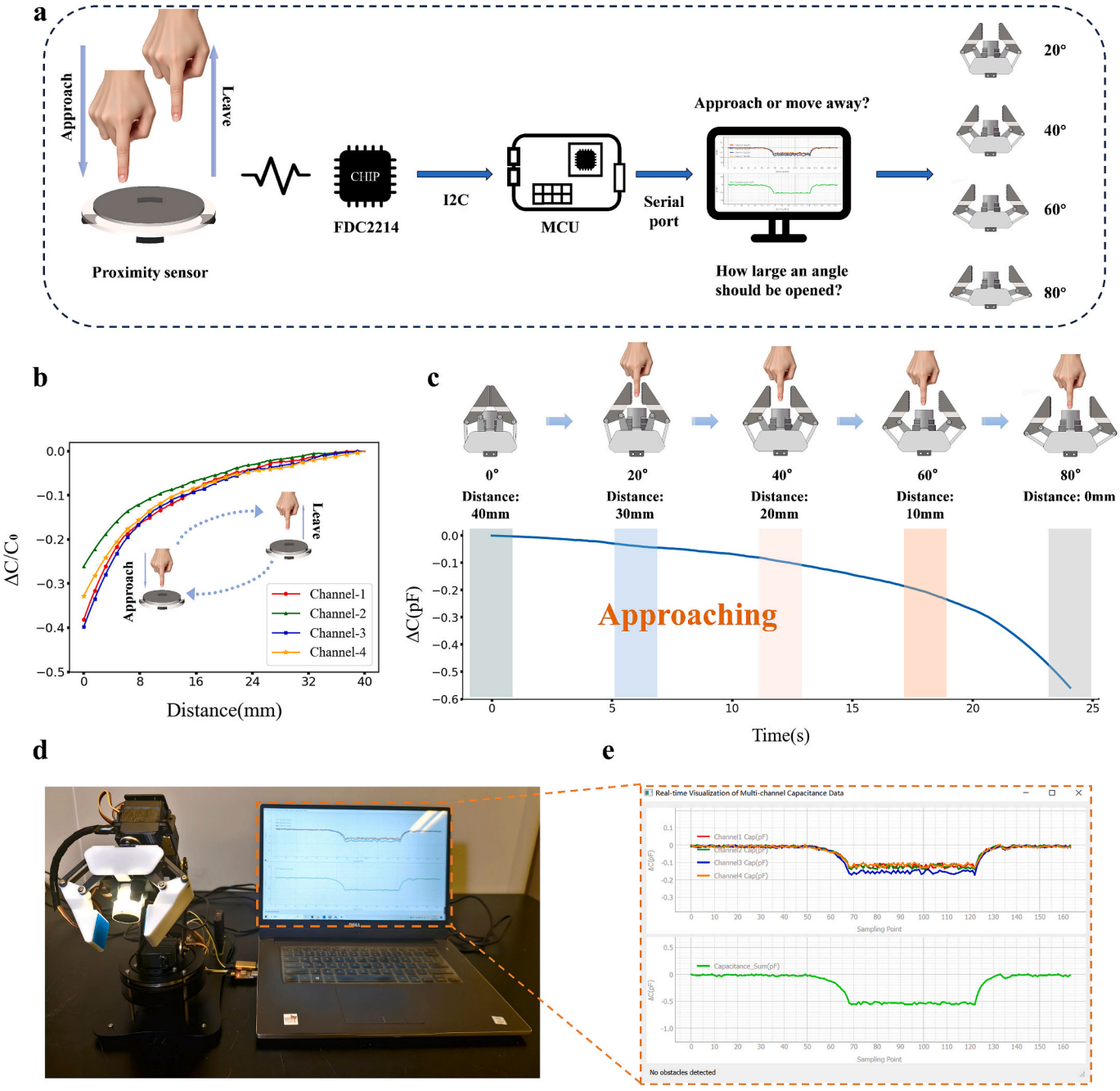

Fig. 5. Active obstacle avoidance system for robotic gripper based on distance sensing function of VTP sensor. (a) Workflow diagram of active obstacle avoidance system; (b) Correspondence between experimenter's finger approach level and capacitance signal of VTP sensor; (c) Mapping relationship between the experimenter's finger approach level and gripper opening angle; (d) Physical device diagram of active obstacle avoidance system; (e) Real-time visualization interface of active obstacle avoidance system.

block. Standardized pressing tests were carried out at preset depths using a tensile testing machine, with reconstruction results shown in Fig. 4e. Experiments demonstrate that under standardized pressing conditions, the 3D geometric shapes of contact surfaces for all three sample types are accurately reconstructed, and reconstruction results clearly reveal the key structural features of object surfaces. This not only verifies the sensor's application potential in precision sensing of object surfaces but also provides reliable experimental evidence for subsequent object recognition tasks based on tactile information.

3.3. Active obstacle avoidance of robotic gripper based on distance sensing

To enhance the safety of human-robot interaction, an active obstacle avoidance system for robotic grippers based on the VTP sensor is designed. Its primary function is to continuously detect the approaching distance of objects in front during gripper operation in real-time. When human body parts (such as fingers) enter the sensing range, active obstacle avoidance is achieved by adjusting the opening angle of the gripper to avoid the risk of pinch injuries [29]. The dynamic process of the gripper opening to avoid obstacles when a finger approaches the robotic arm can be intuitively demonstrated in Movie S1. The system workflow is depicted in Fig. 5a. The proximity sensing unit of VTP sensor continuously collects signals indicating the approach of objects in real-time and converts them into capacitance values, which are sampled by the capacitance acquisition chip and transmitted to the microcontroller via I2C protocol. Subsequently, they are sent to the computer through serial communication. The Python program on the computer performs threshold determination on the capacitance signals and generates control commands for opening or closing the gripper to achieve active obstacle avoidance.

Experimental testing (Fig. 5b) reveals that as the experimenter's finger gradually approaches the interior of the gripper, the relative change in capacitance of the proximity sensing unit of VTP sensor gradually increases. The amplitude of the change is more significant at shorter distances, which provides a solid signal basis for distance quantification detection. To optimize the system's response speed and judgment accuracy, the four-channel capacitance change values are summed and then subjected to threshold determination, establishing a mapping relationship between capacitance change and the opening angle of the gripper (Fig. 5c). When the gripper is in a non-working state, it stays closed. During its operation, it assumes an open state. When the finger approach causes the capacitance change to reach specific ranges, the gripper progressively increases its opening angle. When the distance between the finger and the VTP sensor approaches 0 mm, ΔC_{total} is approximately -0.558 pF, and the gripper opening angle reaches maximum. Fig. 5d depicts the physical device of the active obstacle avoidance system, and Fig. 5e represents the real-time sensing data in the visualization interface, which can dynamically display four-channel capacitance signal curves, summation data, offering intuitive support for system debugging and performance monitoring. This active obstacle avoidance scheme significantly enhances the safety and intelligence level of human-robot collaboration through its pre-contact detection capability and holds important application value in fields with stringent safety requirements such as medical robotics.

3.4. Neural network-based multi-mode recognition of VTP sensor

The VTP sensor proposed in this paper realizes functional integration of near-distance proximity sensing and high-resolution tactile imaging via the integrated design of capacitive proximity sensing unit and visual-tactile sensing unit, providing a fascinating technical solution for enhancing sensing capabilities of robot fingertips. Among these, the capacitive proximity sensing unit employs a four-channel sensing architecture capable of real-time acquisition of capacitance time-series signals throughout the entire process from object approach to contact. The visual-tactile sensing unit captures tactile image information when objects contact through a built-in camera. Thanks to the coordinated working mechanism of the two units: during the object approach phase, different materials exhibit significant differences in capacitance change patterns of the capacitive proximity sensing unit due to variations in properties such as electrical conductivity and dielectric constant [30,31]. During the object contact phase, tactile image features differ for objects with different geometric shapes [32]. This characteristic endows the VTP sensor to have multi-mode object recognition capability. Data fusion methods are generally classified into three categories, namely data-level fusion, feature-level fusion, and decision-level fusion. Data-level fusion directly performs merging, superposition, or weighted averaging on dual-source raw signals; feature-level fusion first extracts features from raw data of different types of sensors separately before fusing the feature vectors; decision-level fusion involves each sensor processing its data independently, followed by fusing the final decision results [33,34]. This study adopts decision-level fusion to process multi-modal information of the VTP sensor: First, the capacitive proximity sensing module collects capacitance data during the object approaching process, which is computed by the processor to obtain material recognition information and temporarily stored. Subsequently, the visual-tactile sensor acquires image data during the contact process, and object recognition information is obtained through computation. After the two types of modal information are processed independently in sequence, multi-dimensional information of objects with specific materials is finally integrated. The feature extraction of physical properties and result output by the algorithm are illustrated in Fig. 6a, and the algorithm processing flow is shown in Fig. 7d.

To systematically evaluate the material recognition performance of the VTP sensor, eight typical materials characterized by smooth surfaces and widespread daily applications were carefully selected as test samples, including copper (Cu), aluminum (Al), iron (Fe), wood, glass, acrylic, nylon, and resin (Fig. S18). There are fundamental disparities in the modulation effects of different materials on the fringe field of the capacitive proximity sensing unit [31], which serves as the crucial physical foundation for achieving non-contact material recognition. Throughout the entire process from object approach to contact with the sensor, conductive materials interact indirectly with the electric field generated by the sensor via electrostatic induction effects. In contrast, insulating materials influence the electric field distribution through polarization effects [30]. Owing to the varying properties such as electrical conductivity and dielectric constant among different materials, the intensity and patterns of their modulation on the electric field show significant differences. These differences are directly reflected in the capacitance value change curves of the VTP sensor. As depicted in Fig. 6b, capacitance time-series curves corresponding to different materials possess distinctive differentiation features. Moreover, the capacitance curves for the same material after multiple cycle tests exhibit excellent consistency and repeatability, verifying the advantageous reliability of the signals. To further verify the feature distinguishability of materials, t-Distributed Stochastic Neighbor Embedding (t-SNE) dimensionality reduction was employed to analyze the capacitance sequence data. The results (Fig. 6c) indicate that the feature data for different materials form well-defined clustering groups in the low-dimensional space, suggesting that the feature differences between materials can be effectively extracted. Based on the aforementioned capacitance sequence data, a multi-layer perceptron (MLP) model was established for material feature extraction and classification (Fig. 6d). The test results, presented via a confusion matrix (Fig. 6e), indicate that the model achieves an average classification accuracy of 94.1% for the eight materials, confirming that the VTP sensor can achieve high-precision material identification by leveraging the material property information embedded in capacitance signals.

To evaluate the VTP sensor's perception and classification capability regarding object surface geometric features, 26 English letters were selected as representative geometric shape samples for testing

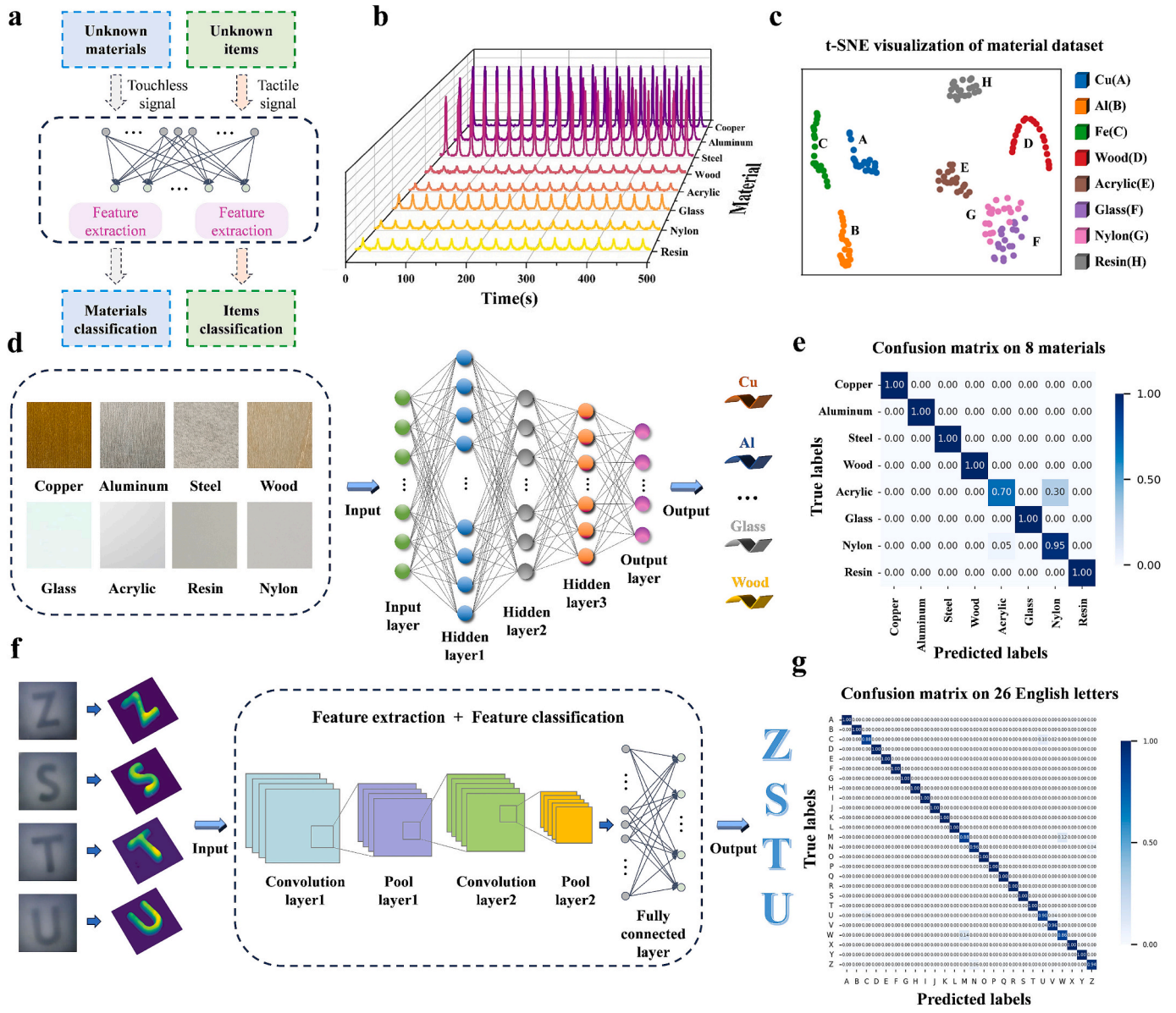

Fig. 6. Neural network-based multi-mode recognition system of VPS sensor. (a) Flowchart of neural network-based multi-mode recognition for identifying unknown material types and objects; (b) Capacitance sequence dataset of 8 materials collected by capacitive proximity sensing unit of VTP sensor (showing one capacitance channel as an example); (c) t-SNE plot of material type classification dataset; (d) Multi-layer perceptron (MLP) model for material classification; (e) Confusion matrix of MLP model material classification results; (f) Convolutional neural network (CNN) model for tactile image classification; (g) Confusion matrix of CNN model's letter classification results.

(Fig. S19). During experiments, operators pressed letter samples almost vertically on the sensor surface. Taking into account the randomness of pressing depth and angle in real-world operations, some tactile images exhibit an uneven distribution of pixel values, leading to local non-uniformity in the 3D contact geometric shape reconstruction. Nevertheless, the overall reconstruction effects still accurately reflect core geometric information of object surfaces (Fig. 6f). Based on tactile images captured by the visual-tactile sensing unit, a tactile image dataset containing 26 sample classes was established. To enhance the model's recognition robustness in complex practical scenarios, data augmentation processing was performed on the dataset, including generalization operations such as angle rotation, position offset, and brightness adjustment, which guarantees stable recognition capability for letter tactile images obtained under different pressing conditions. The augmented dataset was proportionally divided into training and test sets. A convolutional neural network (CNN) model was constructed for tactile image feature extraction and classification (Fig. 6f), with the test results displayed via a confusion matrix (Fig. 6g). The model achieves an average classification accuracy of 97.6% for the 26 English letters, demonstrating that the VTP sensor can effectively leverage the object surface geometric shape information contained in tactile images to achieve high-precision object classification and recognition.

3.5. *Application of multi-mode recognition in intelligent sorting of VTP sensor*

To further validate its application potential of VTP sensor in actual industrial scenarios, the VTP sensor was thoroughly integrated with a robotic arm control system to develop an intelligent sorting system based on multi-mode perception. The system achieves precise recognition of object material and shape, and automated sorting through the coordinated dual-mode sensing capability of the VTP sensor and neural

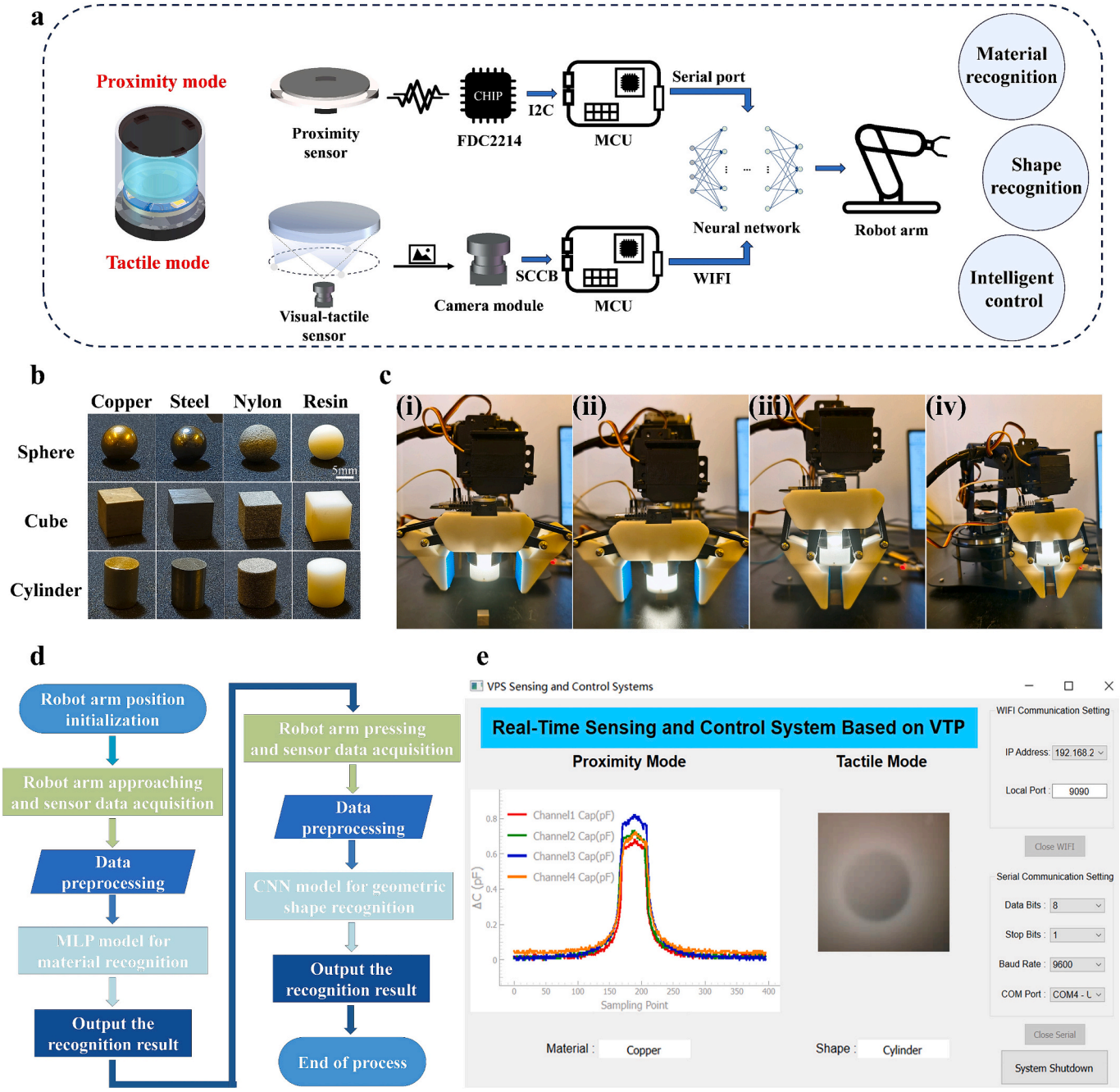

Fig. 7. Application of multi-mode recognition system of VTP sensor in intelligent sorting. (a) Overall workflow diagram of the robotic arm intelligent sorting system; (b) Experimental test samples: combination samples of 4 materials (copper, iron, nylon, resin) and 3 geometric shapes (sphere, cube, cylinder); (c) Demonstration diagram of the intelligent sorting process: (i) Initial positioning of the robotic arm; (ii) Proximity sampling phase; (iii) Contact imaging phase; (iv) Gripping and transfer phase; (d) Core workflow diagram of object recognition program; (e) Graphical user interface (GUI) of intelligent sorting system.

network algorithms, with the overall workflow presented in Fig. 7a. The VTP sensor employs a dual-mode perception mechanism: in the proximity sensing mode, the capacitance sampling chip collects capacitance signals, which are then transmitted to the computer through serial communication via the MCU (microcontroller unit). In the contact sensing mode, the camera of the visual-tactile sensing unit captures tactile images. These images are transmitted to the microcontroller via SCCB bus, and then uploaded to the computer via WIFI communication. Both types of sensing signals are input into pre-trained neural network models on the computer for inference computation, ultimately yielding recognition results of the object's material and shape. The system generates sorting commands based on these results to control the robotic arm to transport objects to the corresponding target areas (The detailed workflow is shown in Fig. S20). It should be noted that in actual industrial scenarios, the target areas can be expanded to specific workstations such as shelves and production lines. The simulation experiment in this study only employs different desktop zones as sorting targets to verify the feasibility of the core system functionality.

Fig. 7d elaborates on the Python program workflow for real-time data acquisition and recognition. Initially, the robotic arm performs initialization operations, moving to the preset initial position to wait for task initiation. Subsequently, the robotic arm approaches the object uniformly from above until it makes contact with the object, and then returns to the initial position along the original path at a uniform speed.

During this process, the capacitive proximity sensing unit of VTP sensor synchronously collects four-channel capacitance time-series data. After preprocessing, these data are input into the trained MLP model to complete object material recognition and output the results. Afterward, the robotic arm approaches the object again at uniform speed and maintains contact. The visual-tactile sensing unit captures tactile images in real-time. After preprocessing operations such as grayscale conversion, noise reduction, and size normalization, these images are input into the pre-trained CNN model for geometric shape recognition, and the shape classification results are output. At this point, the multi-dimensional object recognition process is completed, achieving a closed loop of real-time sensing data acquisition, processing, and recognition.

Notably, single-pass multimodal recognition is technically feasible, and the core optimization strategies can be summarized as follows. First, multi-threaded parallel acquisition is adopted to synchronously collect visual-tactile images and capacitive time-series signals within a single robotic arm movement cycle. These two types of signals are transmitted to the host computer through different development boards and communication protocols, where two parallel data-receiving threads are employed to acquire data from the wired serial port and wireless network separately, and all data are temporarily stored in pre-allocated static memory of a fixed size. Second, feature-level fusion is performed on the dual-modal data after preprocessing; the four sets of one-dimensional capacitive time-series signals and one frame of grayscale image are then spliced into a two-dimensional vector to integrate material features during object approach and morphological features during contact. Third, an end-to-end multimodal fusion model is constructed, which takes the spliced two-dimensional vector as input and directly outputs the joint classification probability of material and shape in a single inference, with the combination of the highest probability selected as the final recognition result. Finally, the recognition results are synchronously transmitted to the UI thread for real-time output of material and shape identification information after model inference.

In the prototype stage of this work, neither single-pass acquisition nor joint recognition is implemented. The proposed sequential two-step workflow is designed in accordance with the bionic perception mechanism of *Tursiops truncatus*, which follows the core logic of non-contact perception first, contact execution later, rather than simply focusing on material and shape recognition. This sequential framework avoids complex parameter matching between the motion speed of the robotic arm and the sampling frequency of multi-sensors. It also realizes complete decoupling of proximity signals and tactile signals, which effectively reduces signal interference and simplifies system debugging. The core objective of the present study is to verify the dual-mode sensing performance of the VTP sensor and the fundamental feasibility of bionic multimodal perception. Moreover, the sequential two-step workflow is more conducive to stable experimental verification and independent analysis of recognition results.

To verify the sorting performance of the system, four typical materials (copper, iron, nylon, resin) were carefully selected and precisely processed into three geometric shapes (sphere, cube, cylinder), thereby constructing 12 combination test samples (Fig. 7b). The specific execution steps of the intelligent sorting system are as follows (Fig. 7c): (i) The robotic arm moves to the preset initial position to accurately complete positioning. Simultaneously, it starts the intelligent sorting system GUI (Fig. 7e), with the interface beginning to display the system's working status and sensing data in real-time; (ii) The gripper descends at a constant speed until the VTP sensor comes into contact the object, and then returns along the original path to the initial position. During this process, collected capacitance data is transmitted to the computer and undergoes material recognition through the MLP model. The recognition results are displayed in real-time on the GUI interface; (iii) The robotic arm approaches the object again at uniform speed and maintains contact. The VTP sensor captures one frame of clear tactile image, which completes shape recognition through the CNN model after preprocessing, with geometric shape information synchronously fed back to the interface; (iv) The system integrates the material and shape recognition results and stores them, generating targeted sorting commands to control the gripper to precisely grasp the object and transport it to the corresponding target area according to the preset rules (The detailed process is shown in Movie S2). Throughout the working process, the system GUI displays four-channel capacitance signal curves, tactile images, recognition results, and robotic arm's working status in real-time, achieving full-process monitoring and traceability of the sorting process (Fig. S21 and Fig. S22).

The intelligent perception system of VTP sensor developed in this study integrates geometric shape feature recognition and material recognition functions, endows mechanical equipment with richer and more precise environmental perception capabilities, effectively enhancing operational reliability and efficiency in complex sorting tasks. Moreover, this system holds important application value and promising promotion prospects in fields such as intelligent workshop automated sorting, medical care precision operations, and human-robot collaboration safety interaction.

4. Conclusion

This study has successfully developed a Visual-tactile-proximity (VTP) sensor that, by incorporating carbon nanotube (CNT) composite materials into its sensitive units, leverages their excellent electrical conductivity, mechanical flexibility, and high specific surface area to enhance tactile sensing and proximity detection performance, thus achieving a synergetic integration of visual-tactile sensing functionality and capacitive proximity sensing functionality. It is capable of simultaneously realizing tactile surface 3D geometric reconstruction, proximity distance sensing, and recognition of object geometric shape and material category. Experimental testing and simulation verification have demonstrated that the VTP sensor possesses excellent sensing performance. In practical applications, we have successfully integrated the VTP sensor into a robotic arm sensing and control system, achieving active obstacle avoidance of robotic grippers based on capacitance signal threshold determination, which effectively enhances the safety of human-robot interaction. The integration of deep learning algorithms has enabled the intelligent recognition of object materials and shapes, with classification accuracies reaching 94.1% and 97.6%, respectively. By integrating multimodal perception capabilities, this sensor provides a technical solution for environmental interaction and precision operations of intelligent robots, with broad application prospects in fields such as human-robot collaboration, intelligent manufacturing, and medical care. Future research will further expand the diversity of data samples, optimize the structures of deep learning models to enhance recognition robustness in complex scenarios, and explore practical applications of the sensor in broader industrial and civilian fields, thus providing continuous technical support for the in-depth development of machine intelligence.

CRediT authorship contribution statement

Hang Yu: Writing – original draft, Methodology, Investigation, Formal analysis, Conceptualization. **Zhao Xu:** Writing – original draft, Visualization, Software, Methodology. **Haowen Chen:** Visualization, Validation, Methodology, Investigation. **Guangchao Wang:** Visualization, Validation, Software, Methodology, Investigation. **Qiang Sun:** Visualization, Validation, Software, Investigation. **Haoran Fu:** Visualization, Investigation. **Huaping Wu:** Validation, Supervision. **Aiping Liu:** Writing – review & editing, Validation, Supervision, Funding acquisition.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This work was supported by the Zhejiang Provincial Natural Science Foundation of China (no. LZ24A020004 and LRG25A020001), the National Natural Science Foundation of China (no. 12272351, 62401509 and 12372168), the Youth Top-notch Talent Project of Zhejiang Ten Thousand Plan of China (no. ZJWR0308010).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.cej.2026.177786.

Data availability

Data will be made available on request.

References

[1] A.C. Abad, A. Ranasinghe, Visuotactile Sensors With Emphasis on GelSight Sensor: A Review, IEEE Sensors J. 20 (2020) 7628–7638, https://doi.org/10.1109/JSEN.2020.2979662.
[2] A. Wilson, H.L. Jiang, W.Z. Lian, W.Z. Yuan, IEEE, Cable Routing and Assembly using Tactile-driven Motion Primitives, in: IEEE Int. Conf. Robot. Autom. (ICRA), London, ENGLAND, 2023, pp. 10408–10414.
[3] N. Zhang, J. Ren, Y. Dong, X. Yang, R. Bian, J. Li, G. Gu, X. Zhu, Soft robotic hand with tactile palm-finger coordination, Nat. Commun. 16 (2025) 2395, https://doi.org/10.1038/s41467-025-57741-6.
[4] J. Zhou, J. Li, H. Jia, K. Yao, S. Jia, J. Li, G. Zhao, C.K. Yiu, Z. Gao, D. Li, B. Zhang, Y. Huang, Q. Zhuang, Y. Yang, X. Huang, M. Wu, Y. Liu, Y. Gao, H. Li, Y. Hu, R. Shi, M. Mukherji, Z. Zheng, X. Yu, Mormyroidea-inspired electronic skin for active non-contact three-dimensional tracking and sensing, Nat. Commun. 15 (2024) 9875, https://doi.org/10.1038/s41467-024-54249-3.
[5] S.H. Wang, W.F. Cai, Y.L. Wu, S.Y. Zhou, W.L. Hu, J. Zhou, Y. Tian, Y. Song, A. P. Liu, Y. Qiu, L.Q. Zhu, H.P. Wu, A bioinspired multimodal nociceptive system enabling graded pressure–damage pain perception for adaptive robotic interaction, Chem. Eng. J. 535 (2026) 175389, https://doi.org/10.1016/j.cej.2026.175389.
[6] T. Liang, Z. Liu, H. Zhang, X. Zhou, Y. Liang, The structure, material and performance of multi-functional tactile sensor and its application in robot field: A review, Mater. Today 86 (2025) 452–481, https://doi.org/10.1016/j.mattod.2025.03.031.
[7] E. Donlon, S. Dong, M. Liu, J. Li, E. Adelson, A. Rodriguez, GelSlim: A High-Resolution, Compact, Robust, and Calibrated Tactile-sensing Finger, Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS) 2018 (1927-1934), https://doi.org/10.1109/IROS.2018.8593661.
[8] S.J. Li, Z.H. Wang, C.S. Wu, X. Li, S. Luo, B. Fang, F.C. Sun, X.P. Zhang, W.B. Ding, When Vision Meets Touch: A Contemporary Review for Visuotactile Sensors From the Signal Processing Perspective, IEEE J. Sel. Top. Signal Process. 18 (3) (2024) 267–287, https://doi.org/10.1109/JSTSP.2024.3416841.
[9] W. Yuan, S. Dong, E.H. Adelson, GelSight: High-Resolution Robot Tactile Sensors for Estimating Geometry and Force, Sensors 17 (2017) 2762, https://doi.org/10.3390/s17122762.
[10] S.X. Zhang, Y.Y. Yang, Y.H. Sun, N.L. Liu, F.C. Sun, B. Fang, Artificial skin based onvisuo-tactile sensing for 3D shape reconstruction: material, method, and evaluation, Adv. Funct. Mater. 35 (1) (2025), https://doi.org/10.1002/adfm.202411686.
[11] Y.H. Sun, N. Cheng, S.X. Zhang, W.Z. Li, L.Y. Yang, S.W. Cui, H.P. Liu, F.C. Sun, J. W. Zhang, D. Guo, W.J. Han, B. Fang, Tactile data generation and applications based on visuo-tactile sensors: A review, Inf. Fusion 121 (2025) 103162, https://doi.org/10.1016/j.inffus.2025.103162.
[12] Y. Dong, J. Ren, Z. Liu, Z. Peng, Z. Yuan, N. Zhang, G. Gu, Look-to-touch: a vision-enhanced proximity and tactile sensor for distance and geometry perception in robotic manipulation, IEEE/ASME Trans. Mechatron. (2025), https://doi.org/10.1109/TMECH.2025.3645239 early access.
[13] J. Yin, P. Shah, N. Kuppuswamy, A. Beaulieu, A. Uttamchandani, A. Castro, J. Pikul, R. Tedrake, Proximity and Visuotactile Point Cloud Fusion for Contact Patches in Extreme Deformation, Proc. IEEE Int. Conf. Robot. Autom. (ICRA) (2025) 16571–16577.
[14] S.L. Mu, S.J. Li, H.F. Zhao, Z.H. Wang, X. Xiao, X. Xiao, Z.A. Lin, Z.W. Song, H. Z. Tang, Q.H. Xu, D.K. Wang, W.W. Lee, C.S. Wu, W.B. Ding, A platypus-inspired electro-mechanosensory finger for remote control and tactile sensing, Nano Energy 116 (2023), https://doi.org/10.1016/j.nanoen.2023.108790.
[15] F.Y. Guan, Y. Xie, H.X. Wu, Y. Meng, Y. Shi, M. Gao, Z.Y. Zhang, S.Y. Chen, Y. Chen, H.P. Wang, Q.B. Pei, Silver Nanowire-Bacterial Cellulose Composite Fiber-Based Sensor for Highly Sensitive Detection of Pressure and Proximity, ACS Nano 14 (11) (2020) 15428–15439, https://doi.org/10.1021/acsnano.0c06063.
[16] J.R. Huang, H.T. Wang, J.A. Li, S.M. Zhang, H. Li, M. Xin, K. Yan, W. Cheng, D. W. He, X.R. Wang, L.J. Pan, Y. Shi, High-Performance Flexible Capacitive Proximity and Pressure Sensors with Spiral Electrodes for Continuous Human-Machine Interaction, ACS Mater. Lett. 4 (11) (2022) 2261–2272, https://doi.org/10.1021/acsmaterialslett.2c00860.
[17] S.Y. Wang, H.Q. Zhen, H.S. Zhao, D.L. Zhang, H.X. Yue, Visual-tactile inference in soft grippers leveraging piezoionic hydrogels and neuromorphic computing, Chem. Eng. J. 522 (2025) 167677, https://doi.org/10.1016/j.cej.2025.167677.
[18] J.P. Wu, H.M. Pang, L. Ding, Y. Wang, X.K. He, Q.A. Shu, S.H. Xuan, X.L. Gong, A lightweight, ultrathin aramid-based flexible sensor using a combined inkjet printing and buckling strategy, Chem. Eng. J. 421 (2021) 129830, https://doi.org/10.1016/j.cej.2021.129830.
[19] S.E. Navarro, S. Muhlbacher-Karrer, H. Alagi, H. Zangl, K. Koyama, B. Hein, C. Duriez, J.R. Smith, S. EscaidaNavarro, Proximity Perception in Human-Centered Robotics: A Survey on Sensing Systems and Applications, IEEE Trans. Robot. 38 (3) (2022) 1599–1620, https://doi.org/10.1109/TRO.2021.3111786.
[20] H.X. Xue, F.C. Sun, Z.R. Zhu, H. Sun, C.X. Pan, L.W. Li, X. Pu, Z.X. Gan, Enhanced morphology perception of vision-based robotic tactile sensors by multi-metal nanofilms, Chem. Eng. J. 505 (2025) 159029, https://doi.org/10.1016/j.cej.2024.159029.
[21] T. Hüttner, L. Von Fersen, L. Miersch, N.U. Czech, G. Dehnhardt, Behavioral and anatomical evidence for electroreception in the bottlenose dolphin (Tursiops truncatus), Anat. Rec. (Hoboken) 305 (3) (2022) 592–608, https://doi.org/10.1002/ar.24773.
[22] T. Hüttner, L. von Fersen, L. Miersch, G. Dehnhardt, Passive electroreception in bottlenose dolphins (*Tursiops truncatus*): implication for micro- and large-scale orientation, J. Exp. Biol. 226 (22) (2023) jeb245845, https://doi.org/10.1242/jeb.245845.
[23] C.Y. Lin, Z.Q. Lin, S.X. Wang, H.Z. Xu, IEEE, DTact: A Vision-Based Tactile Sensor that Measures High-Resolution 3D Geometry Directly from Darkness, Proc. IEEE Int. Conf. Robot. Autom. (ICRA) (2023) 10359–10366.
[24] C.Y. Lin, H. Zhang, J.K. Xu, L. Wu, H.Z. Xu, 9DTact: A Compact Vision-Based Tactile Sensor for Accurate 3D Shape Reconstruction and Generalizable 6D Force Estimation, IEEE Robot. Autom. Lett. 9 (2) (2024) 923–930, https://doi.org/10.1109/LRA.2023.3339397.
[25] P. Fryn, S. Lalik, N. Górska, A. Iwan, M. Marzec, Comparison of the dielectric properties of Ecoflex(R) with L,D-Poly(Lactic Acid) or Polycaprolactone in the presence of SWCN or 5CB, Materials 14 (7) (2021) 1719, https://doi.org/10.3390/ma14071719.
[26] H.S. Niu, H. Li, N. Li, H.K. Niu, Y. Li, S. Gao, G.Z. Shen, Fringing-effect-based capacitive proximity sensors, Adv. Funct. Mater. 34 (51) (2024) 202409820, https://doi.org/10.1002/adfm.202409820.
[27] Y.X. Li, P. Zhang, L.S. Huang, H.L. Sui, S. Wang, C. Ding, A flexible capacitive proximity-tactile dual-mode sensor based on the biomimetic petal-like electrode, Sens. Actuator A: Phys. 379 (2024) 115975, https://doi.org/10.1016/j.sna.2024.115975.
[28] W.K. Yuan, Y. Ding, J.J. Bian, G.F. Wang, General load-depth relations for spherical, conical, and flat-ended cylindrical indentations of soft elastic layers: From ultra-thin-film to half-space, Mech. Mater. 211 (2025) 105502, https://doi.org/10.1016/j.mechmat.2025.105502.
[29] Z. Saleem, F. Gustafsson, E. Furey, M. McAfee, S. Huq, A review of external sensors for human detection in a human robot collaborative environment, J. Intell. Manuf. 36 (4) (2025) 2255–2279, https://doi.org/10.1007/s10845-024-02341-2.
[30] W.J. Liu, F.H. Xiang, D.Q. Mei, Y.C. Wang, A Flexible Dual-Mode Capacitive Sensor for Highly Sensitive Touchless and Tactile Sensing in Human-Machine Interactions, Adv. Mater. Technol. 9 (4) (2024) 202301685, https://doi.org/10.1002/admt.202301685.
[31] G.M. Ma, F.W. Guo, Y. Li, X. Luo, C.Y. Luo, Q.X. Jin, H. Wu, J.L. Fu, M.T. Zhang, Y. Long, A novel 3D-Printed self-healing, touchless, and tactile multifunctional flexible sensor inspired by cutaneous sensory organs, Compos. Commun. 54 (2025) 102287, https://doi.org/10.1016/j.coco.2025.102287.
[32] A. Alspach, K. Hashimoto, N. Kuppuswamy, R. Tedrake, IEEE, Soft-bubble: A highly compliant dense geometry tactile sensor for robot manipulation, Proc. IEEE Int. Conf. Soft Robot. (2019) 597–604.
[33] Z.Y. Qiu, J. Martínez-Sánchez, P. Arias-Sánchez, R. Rashdi, External multi-modal imaging sensor calibration for sensor fusion: a review, Inf. Fusion 97 (2023) 101806, https://doi.org/10.1016/j.inffus.2023.101806.
[34] Q. Tang, J. Liang, F.Q. Zhu, A comparative review on multi-modal sensors fusion based on deep learning, Signal Process. 213 (2023) 109165, https://doi.org/10.1016/j.sigpro.2023.109165.