

# Large-scale Deployment of Vision-based Tactile Sensors on Multi-fingered Grippers

Meng Wang<sup>1\*</sup>, Wanlin Li<sup>1\*</sup>, Hao Liang<sup>1</sup>, Boren Li<sup>1</sup>, Kaspar Althoef<sup>2</sup>, Yao Su<sup>1†</sup>, and Hangxin Liu<sup>1†</sup>

**Abstract**—Vision-based Tactile Sensors (VBTSs) show significant promise in that they can leverage image measurements to provide high-spatial-resolution human-like performance. However, current VBTS designs, typically confined to the fingertips of robotic grippers, prove somewhat inadequate, as many grasping and manipulation tasks require multiple contact points with the object. With an end goal of enabling large-scale, multi-surface tactile sensing via VBTSs, our research (i) develops a synchronized image acquisition system with minimal latency, (ii) proposes a modularized VBTS design for easy integration into finger phalanges, and (iii) devises a zero-shot calibration approach to improve data efficiency in the simultaneous calibration of multiple VBTSs. In validating the system within a miniature 3-fingered robotic gripper equipped with 7 VBTSs we demonstrate improved tactile perception performance by covering the contact surfaces of both gripper fingers and palm. Additionally, we show that our VBTS design can be seamlessly integrated into various end-effector morphologies significantly reducing the data requirements for calibration.

## I. INTRODUCTION

Humans exhibit remarkable adaptability in grasping and manipulation tasks, ranging from power-gripping heavy tools to delicate handling of fragile items under a variety of conditions. Robotic devices have, thus far, fallen short in terms of achieving these human-like levels of adaptability. Given the crucial role of tactile perception in human hands' functionality, as various studies have indicated [1–4], conferring a sense of touch could be the key ingredient. Indeed, the utilization of tactile feedback has significantly advanced robots' manipulation skills in bottle cap opening [5], handling fragile objects [6], object reorientation [7], cable manipulation [8] *etc.* However, unlike humans who use the entirety of the hand to sense and grasp, object manipulation in robots is based on tactile sensing that is principally restricted to its fingertips. Large-scale, multi-surface integration of tactile feedback into multi-fingered grippers remains a pivotal challenge due to the complex articulation of these mechanical structures. Finding an effective solution to this issue could, however, radically improve flexibility and robustness in robotic grasping.

The prevalent method of incorporating large-scale, multi-surface tactile sensing into standard robotic grippers is the *augmentation* approach, whereby a dedicated layer of devices, such as force-sensitive resistors [7], piezoresistive materials [9], strain gauges [10], or e-skin [11, 12], is affixed

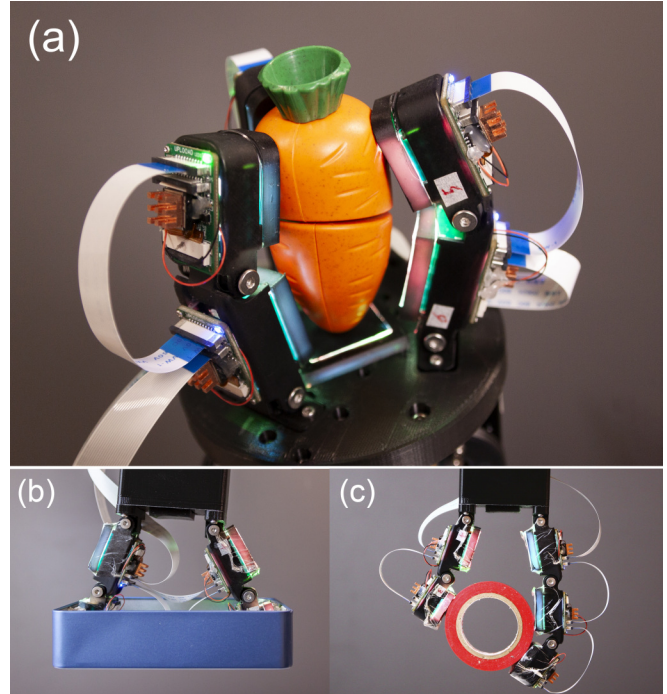


Fig. 1: Large-scale VBTSs deployment on a robotic gripper. (a) Acquiring rich tactile information utilizing the three-fingered robotic gripper. Seven modular VBTSs are deployed into the phalanges of each finger as well as the palm of the gripper. (b)(c) Alternative configurations using modular VBTSs for other tasks.

to the surface of the gripper. However, these setups may well be too rigid to conform to the contours of a multi-fingered gripper, hindering fine manipulative tasks. Additionally, discrepancies between the mounted sensors and the underlying mechanical structure could lead to inaccuracies, reminiscent of the skin-motion artifact observed in human movement. A more effective alternative is the *embedding* approach, in which sensors are integrated into the mechanical structure of the gripper. Typically, these might include high-fidelity sensors like BioTact [13], force transducers, and low-cost VBTSs, such as GelSight [14] or its variants [6, 15–17].

Of all robotic tactile sensor types, VBTSs are deemed the ideal candidates for large-scale deployment on multi-fingered robotic grippers. Using just images, they can acquire pixel-level tactile resolution comparable to human skin - a method significantly easier than constructing sensing arrays like those in e-skin or MEMS transducers - and have contact modules that are elastic and can be easily moulded into different shapes [6, 15–17] to suit various gripper structures and target object properties. However, a key challenge remains,

\*Meng Wang and Wanlin Li contributed equally to this work.

† Corresponding authors.

<sup>1</sup> State Key Laboratory of General Artificial Intelligence, Beijing Institute for General Artificial Intelligence (BIGAI). Emails: {wangmeng, liwanlin, lianghao, liboren, suyao, liuhx}@bigai.ai

<sup>2</sup> Centre for Advanced Robotics @ Queen Mary (ARQ), Queen Mary University of London. Email: k.althoef@qmul.ac.uk.

relating to their bulkiness, despite the potential availability of slimmer versions [17–19]. In addition to this, **data synchronization, easy integration,** and **efficient calibration** are three further difficulties associated with a system involving multiple VBTSs.

In this research, we aim to address these issues and endow a multi-fingered robotic gripper with human-like tactile perception performance by integrating multiple modular VBTSs into each phalange as well as into the palm. To achieve this goal, we first propose a synchronized image acquisition system to mitigate the drawbacks of USB/CSI cameras which are typically used in VBTSs. Then we introduce a modular tactile sensor design that minimizes dimensions, weight, and cabling to ease its integration into finger phalanges and other gripper morphologies. We also develop a zero-shot approach to efficient calibration across multiple sensors. Finally, we validate the entire system creating a 3-fingered robotic gripper in which seven modular tactile sensors operate simultaneously and in which tactile perception is widely achievable across the contact surfaces of both gripper fingers and palm.

#### A. Related Work

**Large-scale Tactile Perception:** The field of large-scale tactile perception has seen significant advancements through innovative sensor technologies [20]. Various techniques, using piezoresistive sensors [9], capacitive sensors [21], piezoelectric sensors [22], magnetic sensors [23], optical sensors [24], *etc.*, have been explored to provide robotic end-effectors with extensive tactile feedback for object assessment via physical contact [25]. Key design parameters such as spatial resolution, sensitivity, wiring complexity, frequency response, flexibility, robustness, and cost are crucial. Balancing these criteria often requires trade-offs in practical applications. While e-skins [11, 12] perform relatively well in the aforementioned categories, they still face challenges such as manufacturing complexity and high cost, as well as having wiring and durability issues.

**GelSight Variants:** VBTSs demonstrate advantages over alternatives in terms of high spatial resolution [14, 26], compliance [27], durability [19], low-cost [28, 29] and wireless [6] characteristics, and, in recent years, several attempts have been made to deploy multiple such VBTSs in robotic end-effectors. For instance, Padmanabha *et al.* [30] utilize three micro-cameras to realize a multi-directional high-resolution tactile sensor; Sferrazza *et al.* [31] describe a multi-camera tactile sensor that shows promise for vision-based robotic skins; Liu *et al.* [32] describe a soft endoskeleton three-fingered robot hand that extracts rich tactile data during grasping; Wilson *et al.* [33] present a two-fingered robot gripper with multiple GelSight sensors that can determine the surface topology of the object; She *et al.* [34] describe a vision-based proprioceptive and tactile sensing finger for soft robots; Zhao *et al.* [35] present a vision-based three-fingered robotic hand that obtains rich tactile signal during grasping; Romero *et al.* [36] demonstrate rounded GelSight fingertips that can be installed on an

TABLE I: The current large-scale vision-based tactile systems.

Name	Maximum No. of sensors working simultaneously	No. of sensors equipped
Omni tact [30]	3	3
Multi-camera Tactile Sensor [31]	4	4
GelSight EndoFlex [32]	4	6
Fully Actuated Robotic Hand With Multiple GelSight [32]	4	4
Vision-based Exoskeleton-covered Soft Finger [34]	4	4
GelSight Svelte Hand [35]	3	3
Soft, Round, and High-Resolution Soft Finger [36]	4	4

Allegro hand to capture contact surfaces in high-resolution. However, as summarized in Tab. I, these attempts present the simultaneous operation of no more than four sensors and lack a synchronizing mechanism to support a large increase in the number of sensors. With this in mind, we try to enhance the design of VBTS to facilitate their large-scale deployment in robotic grippers. A novel three-fingered robotic gripper that incorporates seven VBTSs is presented, in which the entire inner surface of both fingers and palm are imbued with tactile perception, and in which all VBTSs operate simultaneously.

#### B. Overview

We organize the remainder of the paper as follows. Sec. II outlines the hardware design of the proposed modular tactile sensor. Sec. III describes the zero-shot calibration method while Sec. IV presents its deployment on a three-fingered robotic gripper. We discuss and conclude the paper in Secs. V and VI.

## II. MODULAR TACTILE SENSOR DESIGN

### A. Synchronized Image Acquisition System

Traditional GelSight sensors, which utilize USB or CSI cameras to capture vision-based tactile information, pose challenges for deployment across multiple sensing planes

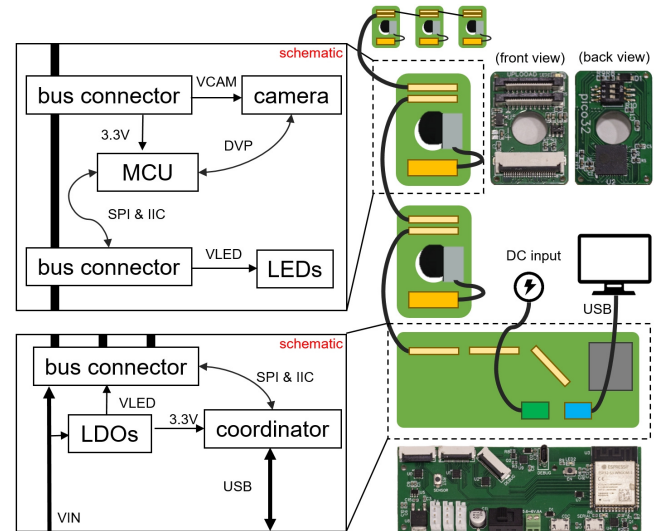


Fig. 2: **Synchronized image acquisition system.** Multiple sensor boards are cascade-connected to the hub board through FFC cables. The hub board provides the power supply for both the sensor boards and the motors, and handles the communication between the acquisition system and the host PC.

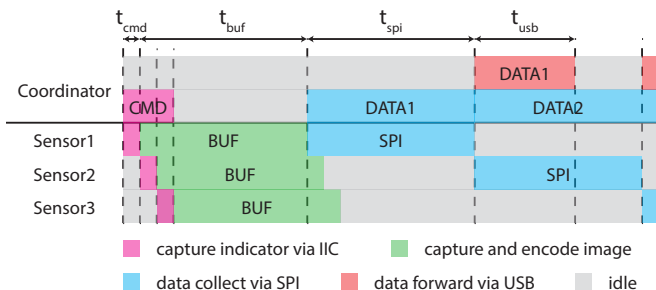


Fig. 3: **Sequence diagram of the acquisition process.** The system is designed to minimize synchronization error, and thus provide a high frame rate with limited latency.

due to their large dimensions and separate wiring. Moreover, these sensors often lack temporal synchronization in image acquisition, a problem that intensifies as the number of sensors increases. To address these drawbacks, we propose a modular image acquisition system specifically designed to incorporate multiple image sensors within the confined spaces of finger-like actuators.

The acquisition system (see Fig. 2) comprises a hub board and multiple sensor boards that can be cascade-connected via FFC cables or customized FPCs. Each sensor board is equipped with an MCU and communicates with a miniature camera through a DVP port. The hub board provides assorted connectors for linking sensor boards, motors, power supply, and the host computer. It also features a core coordinator, which can be either a USB multifunction converter IC (e.g., CH347) or an RTOS-supported MCU for greater flexibility. The coordinator manages the communication bus, which includes an SPI bus for image acquisition and an IIC bus for sensor control. Power modules like LDOs are also placed on the hub board to reduce heat of sensors.

**Temporal Synchronization Design:** To enhance synchronization accuracy, we implement an additional control signal that dictates the precise moment for image capture by the sensors. Operating on a Fast-mode I2C bus at a frequency of 400 kHz, the time required for each sensor to capture an image is remarkably less than  $30 \mu s$ , ensuring high synchronization.

**Theoretical Performance:** Image acquisition commences immediately upon readiness of the first sensor. The coordinator sequentially collects images from the sensors via the SPI bus and forwards them to the host computer via USB. The system performance heavily depends on the number of sensors  $N$ , and the required image frame size  $S$ . Given that the USB bitrate typically surpasses the SPI bitrate, the system's overall frame rate is primarily determined by the latter. To optimize frame rate and minimize latency, the SPI bus operates at a 40 MHz clock frequency in DMA mode. With a common frame size (320x240, JPEG) of  $S = 20$  KB, the transmission time for every image is  $t_{spi} = 4$  ms, while the initial image buffering time  $t_{buf}$  is approximately 1 ms. As such, the frame rate is computed as  $f = 1/(Nt_{spi} + t_{buf})$ , with the maximum latency being  $latency = Nt_{spi} + t_{buf} + t_{usb}$ . The maximum synchronization error is calculated as  $error_{sync} = 30 N\mu s$ . This synchronization timeline is illustrated in Fig. 3.

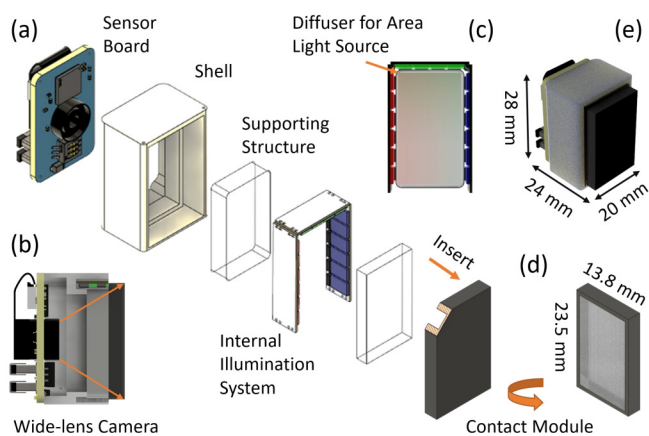


Fig. 4: **Design and fabrication of the modular vision-based tactile sensor.** (a) The structure of the VBTS consists of a modular image acquisition board, a wide-lens camera, an illumination system, a supporting structure, a contact module, and a shell. (b) The view of the wide-lens camera covers the whole surface of the contact module. (c) The internal area lighting system utilizes side illumination towards the supporting structure for shade accentuation and compactness. (d) The elastomer serves as the contact module to provide super-high spatial resolution. (e) The overall dimensions of each sensor are minimized to  $28 \times 20 \times 19$  mm.

## B. Modular Tactile Sensor Design & Fabrication

The dimensions of VBTSs are primarily determined by the image acquisition system. Leveraging the modular system presented above, we can achieve a compact design. As illustrated in Fig. 4(a), the dimensions of the sensor are minimized to  $28 \times 20 \times 19$  mm and its weight to  $20 g$  - figures comparable to those of a human phalanx. The proposed sensor comprises a modular image acquisition board, a wide-lens camera, an internal illumination system, a supporting structure, a contact module, and a shell, as depicted in Fig. 4(b). The fabrication process of each component is detailed below.

**Wide-lens Camera:** As illustrated in Fig. 4(c), we used an OmniVision OV2640 CMOS camera module with a 160-degree viewing lens to provide an open view of 2 megapixels. Apart from camera calibration to counter wide-lens distortion, parameters such as exposure and white balance are manually configured to align with the internal illumination system and ensure optimal sensor performance.

**Modular Image Acquisition Board:** The sensor board accommodates the camera lens in a central aperture, resulting in a compact assembly. An onboard MCU captures and stores images, while two external connectors facilitate bus cascading. In addition, separate power is supplied to the illumination system to provide stable lighting conditions.

**Internal Illumination System:** Amongst several illumination configurations in the design of VBTSs, illumination towards the supporting structure is generally used in flat, compact GelSight sensors [17], operating under the principle of grazing lighting that accentuates surface texture by exaggerating its shades. To achieve this effect, we use multiple flat-surface 3528 SMD LEDs with a light diffuser, creating an area surface lighting source that directly illuminates the side of the supporting structure, as depicted in Fig. 4(d).

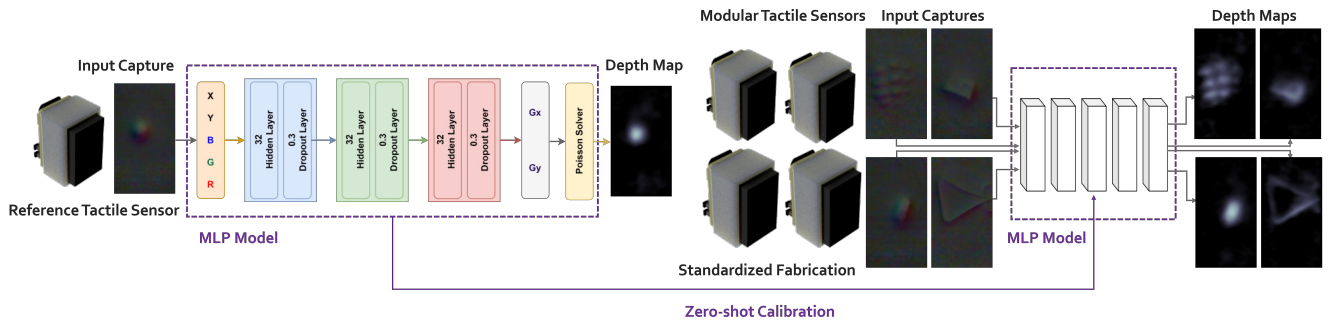


Fig. 5: **The zero-shot calibration process for large-scale deployment of VBTSS.** To jointly calibrate the VBTSS that share the similar configuration, a multiple-layer perception (MLP) neural network is utilized to learn the mapping between pixel intensity within RGB color channels, position information, and gradients in the  $x$  and  $y$  directions.

Therefore, any deformation of the contact module leads to a significant change in the optical path, visible to the camera.

**Supporting Structure:** The supporting structure is in direct contact with the underside of the soft contact module and facilitates the camera’s observation of the deformations on the other side of the module. Rigidity and transparency are therefore key requirements. For ease of self-crafting, optional materials include acrylics (sheet, tube, shells), epoxy resin (reactive prepolymers), and clear resin (Stereolithography 3D printing). Here, we chose a 5 mm thickness acrylic sheet (in accordance with the LED’s dimension) and customized its shape with a laser cutter.

**Contact Module:** VBTSS typically have elastomers as their contact modules, which can be handcrafted [14] or 3D-printed [26]. Handcrafted elastomers offer modular fabrication with finer surfaces, enabling higher spatial resolution. In contrast, 3D printing typically generates dot patterns. In this study, we utilized the handcrafted approach as per the Gelsight sensor in [14], in which the contact module comprises a transparent silicone base and a coating layer with a Lambertian surface. The transparent silicone base is made of Part A&B and silicone thinner Slacker in a weighted ratio of 1 : 1 : 3 which produces a moderate sensing medium. The thickness of the base is fabricated at 3 mm. We airbrushed the diluted mixture of spherical aluminum powder, silicone solvent, and Smooth-On NOVOCSS Matte (weighted ratio of 1 : 10 : 30) onto the base’s upperside to form a reflective coating layer with a Lambertian surface. Finally, we added a thin protective layer made of mill-resistant matte oil to increase the sensor’s durability. The final contact module is depicted in Fig. 4(e)

**Shell:** The shell is essential for securing each component within the sensor while preventing external environmental light from affecting the internal capture system. To achieve this, we utilized Stereolithography (SLA) 3D printing technology to produce the shell from lightweight black resin material. The manufactured shell displays good rigidity and toughness, while effectively blocking ambient light.

### III. MODULAR SENSOR CALIBRATION

In this section, we describe a zero-shot calibration approach to the large-scale deployment of the proposed VBTSS, which have identical dimensions and setups on each phalange of a robotic multi-fingered gripper. We then present

a performance comparison between the proposed calibration approach and individual calibration for multiple sensors.

The principle of GelSight-like tactile sensors is based on photometric stereo which maps the captured pixel intensity to the change in surface gradient [14]. A fast Poisson solver is then applied to integrate gradients to obtain the depth map. Due to its modular design, each sensor module shares almost the same configuration, including identical dimensions, illuminating systems, camera settings, and contact module composite. Though few variations might introduced during fabrication and assembly, it still makes zero-shot calibration feasible with fine-tuning techniques.

In this study, we employ a multiple-layer perception (MLP) neural network to learn the mapping between pixel intensity within three color channels, position information, and gradients in the  $x$  and  $y$  directions, as depicted in Fig. 5. The network is composed of three hidden layers and three dropout layers with the  $Tanh$  activation function. The training dataset for a single sensor consists of 50 captures, and the testing dataset comprises 5 captures. We conducted trials using both the raw capture and the differential capture (between current frame and reference frame) as inputs to the model. When transitioning to another modular VBTSS, we initially fine-tune a single difference capture to achieve alignment with the reference sensor, by applying a constant offset to each color channel. Then we directly apply the model to achieve automatic calibration.

Tab. II presents quantitative results for the reconstruction using different approaches: individual calibration and zero-shot calibration. As can be seen, utilizing the difference input for zero-shot calibration results in errors closer to those of the individual calibration approach as compared to the raw input, while significantly reducing the required dataset size, as well as the calibration time. This proves the effectiveness of the proposed zero-shot calibration approach for rapid large-scale deployment of VBTSS.

TABLE II: **Evaluation of proposed calibration approaches on three modular tactile sensors using look-up table and MLP**

Groups	Calibration Approaches				
	Individual Lookup Table	Individual Raw Input	Individual Diff Input	Zero-shot Raw Input	Zero-shot Diff Input
Data Collected	150	150	150	50	50
$G_x$ error (MAE)	0.041	0.034	0.028	0.042	0.034
$G_y$ error (MAE)	0.048	0.043	0.034	0.051	0.039

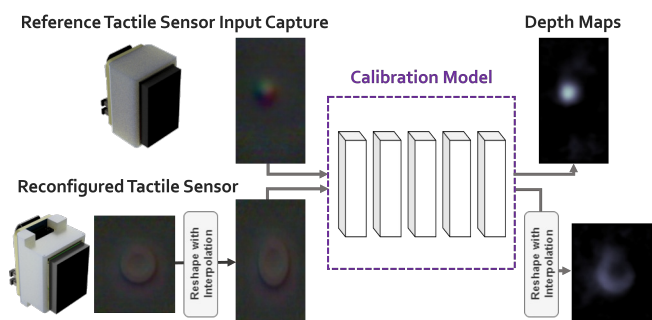


Fig. 6: **The cross-configuration test of the proposed zero-shot calibration model.** Applying the same calibration model on other VBTs which have non-identical configurations, the model still offers high-fidelity depth map reconstruction performance.

Additionally, we conducted an experiment on VBTs with similar but non-identical configurations. The qualitative results, as depicted in Fig. 6, demonstrate that despite some differences in size between the two sensors, the model can still reconstruct the depth map in high-fidelity taking advantage of the proposed zero-shot calibration method, further underscoring its robustness.

#### IV. REAL DEPLOYMENT ON GELGRIPPER

##### A. GelGripper Design

The modular design enables large-scale deployment of the VBTs. Here, we present the deployment details on an end-effector named GelGripper, integrating seven of the proposed VBTs. As illustrated in Fig. 7, the GelGripper is a cable-driven three-fingered robotic gripper imbued with tactile sensing capabilities across its inner surfaces, including its palm. Each rigid finger is composed of a pair of embedded modular tactile sensors with springs to form the joints. The three fingers are symmetrically positioned around the palm base, effectively with 110 degree gaps between them, while an additional tactile sensor is located in the palm's center. The gripper is able to support seven modular sensors working simultaneously to provide large-scale tactile feedback. Cable-driven actuation is applied to maximize the sensing region available and minimize the weight of the finger section. Three Dynamixel XM430-W350-T servo motors are fitted beneath the palm base to drive the cables and achieve flexion through the finger shells, while finger extension is driven by the spring at each joint.

##### B. Grasping with GelGripper

To evaluate the performance of the proposed GelGripper in grasp and manipulation tasks, we used it to replace the end-effector of a Franka Emika Panda 7-Degree-of-Freedoms (DoFs) robot arm with a 3D-printed rigid adapter. As shown in Fig. 8(a), the sensors can capture subtle changes in the contact imprint in very high resolution, enabling gentle handling of deformable objects. Furthermore, Fig. 8(b) illustrates a control test in which a software-simulated synchronization error is introduced to one of the sensors. Although the system attempts to achieve a similarly gentle grip, the actual gripping force exceeds the intended level at the point at which the

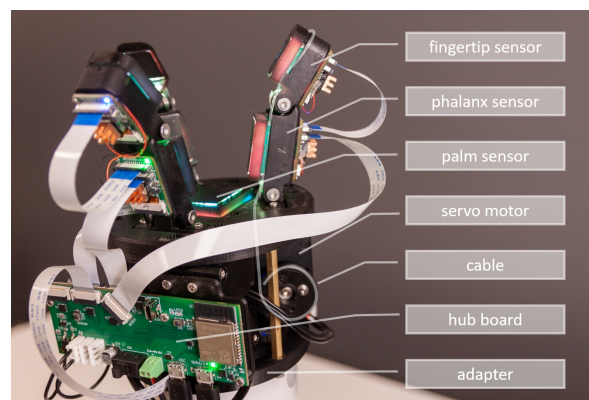


Fig. 7: **The mechanical design of the GelGripper.** Three different types of sensing structures for the fingertip, middle phalanx, and palm are all incorporated within the same sensing hardware. Each finger is actuated by a servo motor through the cable-driven mechanism. The compact design of GelGripper enables seamless integration with robot arms.

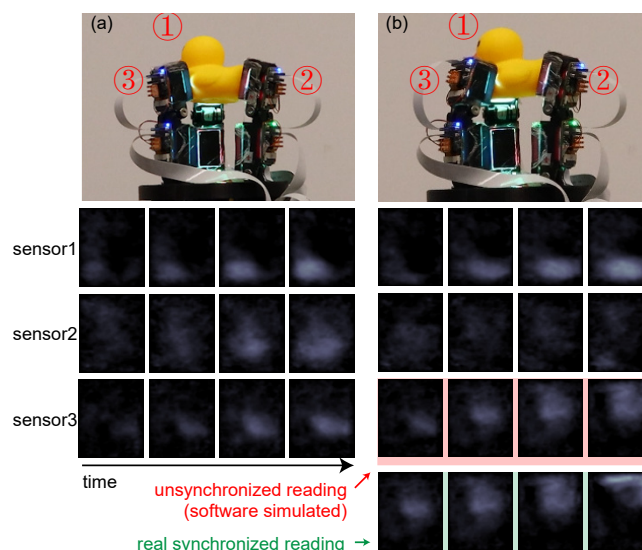


Fig. 8: **Synchronization performance study of VBTs in gripping.** (a) Gentle handling of a rubber duck achieved with synchronized tactile data. (b) The undesired force that exceeds the threshold appears with unsynchronized tactile data.

gripper stops, demonstrating the necessity of tactile data synchronization between multiple sensors.

The GelGripper demonstrates its ability to securely grasp everyday objects of varying shapes and levels of hardness. Fig. 9(a) illustrates how the GelGripper acquires real-time tactile feedback information from its fingers and palm during these grasping interactions. The GelGripper can also manipulate tools to conduct daily tasks, as demonstrated in Fig. 9(b), which illustrates the process of operating an electric screwdriver to tighten a bolt. The entire procedure can be split into five steps: ① at rest; ② secure the body of the screwdriver with the fingers and palm; ③ press the trigger of the screwdriver with one of the fingertips until it starts; ④ wait for the bolt to become tightened, and ⑤ then release the trigger. Fig. 9(b4) illustrates the deformation signal of each sensor throughout the entire procedure.

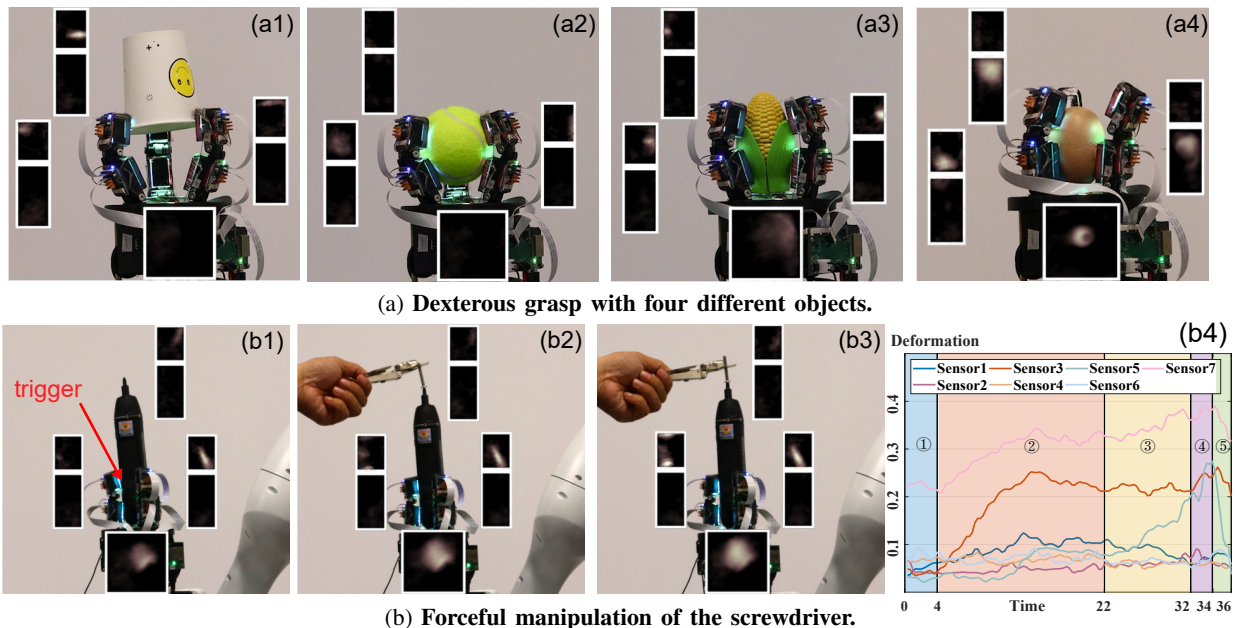


Fig. 9: **Grasp and manipulation with the GelGripper.** (a) Four objects: paper cup (a1), tennis ball (a2), toy corn (a3), egg (a4) are grasped by the GelGripper. (b) The screwdriver is forcefully manipulated by the GelGripper to tighten the bolt with the deformation signal of each sensor plotted in (b4). The whole process is divided into five stages: ① rest status; ② gripping the electric screwdriver body; ③ pressing (b1) and starting the trigger (b2); ④ tightening the bolt (b3); ⑤ releasing the trigger. The imprints of seven VBTSs are also presented in the white boxes within each sub-figure.

## V. DISCUSSION

**Synchronized Image Acquisition:** In contrast to simple parallel grippers, multi-fingered grippers require several sensors operating concurrently to achieve adaptability in grasping. As the number of sensors increases and the frame rate decreases, synchronization becomes essential, to ensure reliable results from subsequent image processing, regardless of fluctuations in frame rate and latency. We developed a prototype in which all sensors were wirelessly connected to facilitate integration with the gripper. However, the prototype struggled to achieve satisfactory synchronization, rendering it sensitive to transmission fluctuations and resulting in poor performance in real-world tasks.

**Data Transmission and Processing:** For data acquisition, judicious hardware and software design of the communication bus and protocol is essential to achieve optimal performance. Although the SPI bus can support clock frequencies up to 40 MHz, the actual speed may be impacted by various factors, such as electromagnetic interference (EMI) and circuit routing. Adding more sensors can also lead to a reduction in transmission speed. Utilizing a self-designed flexible printed circuit with electromagnetic shielding, along with Low Voltage Differential Signalling (LVDS) routing, can significantly enhance performance. Additionally, implementing a COBS-based packeting protocol can improve the efficiency and reliability of data transmission. For data processing, we use multi-threading in Python to enable concurrent computing of tactile data from multiple sources within a quick response time. Different threads share the same memory space, alleviating the burden of computation-intensive executions and resource management, especially when handling multiple sources. Meanwhile, leveraging neu-

ral networks on the GPU helps to offload the CPU resources and mitigate issues such as race conditions during multi-processing.

**Integration:** While VBTSs provide high spatial resolution and rich information, their bulkiness poses a challenge in relation to large-scale deployment. Our system tries to reduce the dimensions and weight of modular tactile sensors by leveraging customized image acquisition boards, modern micro-cameras, and lightweight external and internal supporting designs and materials. Additionally, by employing a cable-driven mechanism we also effectively reduce the overall weight and the dimensions of each finger for real-world grasping. This is notable in the case of multiple phalanges, as an alternative to installing motors at their joints.

**Calibration:** In addition to addressing physical deployment challenges, we have streamlined the software complexity to simplify the calibration process, which can be laborious in real-world scenarios. By employing a modular sensor design, we have implemented a zero-shot calibration method that efficiently reduces calibration time and resources while maintaining high fidelity in tactile data collection.

## VI. CONCLUSION & FUTURE WORK

This paper presented a systematic approach to overcome the difficulties inherent in the large-scale deployment of VBTSs, which principally are data synchronization, ease of integration, and efficient calibration. A synchronized image acquisition system was first innovated for latency minimization. Based on that, a compact modular VBTS was then designed for ease of integration. Finally, a zero-shot calibration approach was proposed to avoid individual

calibration, though without sacrificing accuracy. For demonstration purposes, the overall system was deployed on a three-fingered robotic gripper with seven VBTSs working simultaneously. By integrating the gripper into a robotic arm, the performance of the VBTSs system was verified in both grasp and manipulation tasks.

**Future work** will focus on (i) improving the time efficiency and minimizing the power consumption of the synchronized image acquisition system; (ii) developing a realistic simulator for VBTSs, to further facilitate their efficient calibration under varying conditions; (iii) generalizing this systematic approach to other vision-based tactile sensors for large-scale deployment; (iv) extending demonstrative deployment onto an anthropomorphic 5-fingered robotic hand with more than 15 VBTSs integrated.

**Acknowledgement:** We thank Mr. Mish Toszeghi from QMUL for his meticulous proofreading of our manuscript. This work was supported in part by the National Natural Science Foundation of China (No. 62376031).

#### REFERENCES

- [1] R. Johansson and J. Flanagan, "6.05-tactile sensory control of object manipulation in humans," *Senses Compr. Ref. Acad. Press NY NY*, vol. 67, p. 86, 2008.
- [2] P. Jenmalm, I. Birznieks, A. W. Goodwin, and R. S. Johansson, "Influence of object shape on responses of human tactile afferents under conditions characteristic of manipulation," *European journal of neuroscience*, vol. 18, no. 1, pp. 164–176, 2003.
- [3] T. Schneider and J. Hermsdörfer, "Anticipation in object manipulation: behavioral and neural correlates," *Progress in Motor Control: Theories and Translations*, pp. 173–194, 2016.
- [4] J. R. Flanagan, M. C. Bowman, and R. S. Johansson, "Control strategies in object manipulation tasks," *Current opinion in neurobiology*, vol. 16, no. 6, pp. 650–659, 2006.
- [5] M. Edmonds, F. Gao, H. Liu, X. Xie, S. Qi, B. Rothrock, Y. Zhu, Y. N. Wu, H. Lu, and S.-C. Zhu, "A tale of two explanations: Enhancing human trust by explaining robot behavior," *Science Robotics*, vol. 4, no. 37, 2019.
- [6] W. Li, M. Wang, J. Li, Y. Su, D. K. Jha, X. Qian, K. Althoefer, and H. Liu, "L<sup>3</sup> f-touch: A wireless gelsight with decoupled tactile and three-axis force sensing," *IEEE Robotics and Automation Letters (RA-L)*, vol. 8, no. 8, pp. 5148–5155, 2023.
- [7] Z.-H. Yin, B. Huang, Y. Qin, Q. Chen, and X. Wang, "Rotating without seeing: Towards in-hand dexterity through touch," in *Proceedings of Robotics: Science and Systems (RSS)*, 2023.
- [8] Y. She, S. Wang, S. Dong, N. Sunil, A. Rodriguez, and E. Adelson, "Cable manipulation with a tactile-reactive gripper," *International Journal of Robotics Research (IJRR)*, vol. 40, no. 12-14, pp. 1385–1401, 2021.
- [9] H. Liu, X. Xie, M. Millar, M. Edmonds, F. Gao, Y. Zhu, V. J. Santos, B. Rothrock, and S.-C. Zhu, "A glove-based system for studying hand-object manipulation via joint pose and force sensing," in *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2017.
- [10] "Ati multi-axis force/torque sensors family." <https://www.ati-ia.com/products/ft/sensors.aspx>.
- [11] J. Park, M. Kim, Y. Lee, H. S. Lee, and H. Ko, "Fingertip skin-inspired microstructured ferroelectric skins discriminate static/dynamic pressure and temperature stimuli," *Science advances*, vol. 1, no. 9, p. e1500661, 2015.
- [12] Y. Yu, J. Li, S. A. Solomon, J. Min, J. Tu, W. Guo, C. Xu, Y. Song, and W. Gao, "All-printed soft human-machine interface for robotic physicochemical sensing," *Science Robotics*, vol. 7, no. 67, p. eabn0495, 2022.
- [13] J. A. Fishel and G. E. Loeb, "Sensing tactile microvibrations with the biotac—comparison with human sensitivity," in *IEEE RAS & EMBS international conference on biomedical robotics and biomechatronics (BioRob)*, 2012.
- [14] W. Yuan, S. Dong, and E. H. Adelson, "Gelsight: High-resolution robot tactile sensors for estimating geometry and force," *Sensors*, vol. 17, no. 12, p. 2762, 2017.
- [15] E. Donlon, S. Dong, M. Liu, J. Li, E. Adelson, and A. Rodriguez, "Gelslim: A high-resolution, compact, robust, and calibrated tactile-sensing finger," in *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2018.
- [16] W. Li, A. Alomainy, I. Vitanov, Y. Noh, P. Qi, and K. Althoefer, "F-touch sensor: Concurrent geometry perception and multi-axis force measurement," *Sensors*, vol. 21, no. 4, pp. 4300–4309, 2020.
- [17] S. Wang, Y. She, B. Romero, and E. Adelson, "Gelsight wedge: Measuring high-resolution 3d contact geometry with a compact robot finger," in *IEEE International Conference on Robotics and Automation (ICRA)*, 2021.
- [18] D. Ma, E. Donlon, S. Dong, and A. Rodriguez, "Dense tactile force estimation using gelslim and inverse fem," in *IEEE International Conference on Robotics and Automation (ICRA)*, 2019.
- [19] M. Lambeta, P.-W. Chou, S. Tian, B. Yang, B. Maloon, V. R. Most, D. Stroud, R. Santos, A. Byagowi, G. Kammerer, 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 (RA-L)*, vol. 5, no. 3, pp. 3838–3845, 2020.
- [20] M. L. Hammock, A. Chortos, B. C.-K. Tee, J. B.-H. Tok, and Z. Bao, "25th anniversary article: the evolution of electronic skin (e-skin): a brief history, design considerations, and recent progress," *Advanced materials*, vol. 25, no. 42, pp. 5997–6038, 2013.
- [21] B. Ji, Q. Zhou, M. Lei, S. Ding, Q. Song, Y. Gao, S. Li, Y. Xu, Y. Zhou, and B. Zhou, "Gradient architecture-enabled capacitive tactile sensor with high sensitivity and ultrabroad linearity range," *Small*, vol. 17, no. 43, p. 2103312, 2021.
- [22] W. Lin, B. Wang, G. Peng, Y. Shan, H. Hu, and Z. Yang, "Skin-inspired piezoelectric tactile sensor array with crosstalk-free row+column electrodes for spatiotemporally distinguishing diverse stimuli," *Advanced Science*, vol. 8, no. 3, p. 2002817, 2021.
- [23] Y. Yan, Z. Hu, Z. Yang, W. Yuan, C. Song, J. Pan, and Y. Shen, "Soft magnetic skin for super-resolution tactile sensing with force self-decoupling," *Science Robotics*, vol. 6, no. 51, p. eabc8801, 2021.
- [24] G. Zhang, Y. Du, H. Yu, and M. Y. Wang, "Deltact: A vision-based tactile sensor using a dense color pattern," *IEEE Robotics and Automation Letters (RA-L)*, vol. 7, no. 4, pp. 10778–10785, 2022.
- [25] K. Ahanat, A. Juan, and P. Veronique, "Tactile sensing in dexterous robot hands-review," *Robotics and Autonomous Systems*, vol. 74, pp. 195–220, 2015.
- [26] B. Ward-Cherrier, N. Pestell, L. Cramphorn, B. Winstone, M. E. Giannaccini, J. Rossiter, and N. F. Lepora, "The tactip family: Soft optical tactile sensors with 3d-printed biomimetic morphologies," *Soft Robotics*, vol. 5, no. 2, pp. 216–227, 2018.
- [27] A. Alspach, K. Hashimoto, N. Kuppusswamy, and R. Tedrake, "Soft-bubble: A highly compliant dense geometry tactile sensor for robot manipulation," in *IEEE International Conference on Soft Robotics (RoboSoft)*, 2019.
- [28] A. C. Abad and A. Ranasinghe, "Low-cost gelsight with uv markings: Feature extraction of objects using alexnet and optical flow without 3d image reconstruction," in *IEEE International Conference on Robotics and Automation (ICRA)*, 2020.
- [29] C. Lin, H. Zhang, J. Xu, L. Wu, and H. Xu, "9dtact: A compact vision-based tactile sensor for accurate 3d shape reconstruction and generalizable 6d force estimation," *IEEE Robotics and Automation Letters (RA-L)*, 2023.
- [30] A. Padmanabha, F. Ebert, S. Tian, R. Calandra, C. Finn, and S. Levine, "Omniact: A multi-directional high-resolution touch sensor," in *IEEE International Conference on Robotics and Automation (ICRA)*, 2020.
- [31] C. Trueeb, C. Sferrazza, and R. D'Andrea, "Towards vision-based robotic skins: a data-driven, multi-camera tactile sensor," in *IEEE International Conference on Soft Robotics (RoboSoft)*, 2020.
- [32] S. Q. Liu, L. Z. Yañez, and E. H. Adelson, "Gelsight endoflex: A soft endoskeleton hand with continuous high-resolution tactile sensing," in *IEEE International Conference on Soft Robotics (RoboSoft)*, 2023.
- [33] A. Wilson, S. Wang, B. Romero, and E. Adelson, "Design of a fully actuated robotic hand with multiple gelsight tactile sensors," *arXiv preprint arXiv:2002.02474*, 2020.
- [34] Y. She, S. Q. Liu, P. Yu, and E. Adelson, "Exoskeleton-covered soft finger with vision-based proprioception and tactile sensing," in *IEEE International Conference on Robotics and Automation (ICRA)*, 2020.
- [35] J. Zhao and E. H. Adelson, "Gelsight svelte hand: A three-finger, two-dof, tactile-rich, low-cost robot hand for dexterous manipulation," *arXiv preprint arXiv:2309.10886*, 2023.
- [36] B. Romero, F. Veiga, and E. Adelson, "Soft, round, high resolution tactile fingertip sensors for dexterous robotic manipulation," in *IEEE International Conference on Robotics and Automation (ICRA)*, 2020.