SandWorm: Event-based Visuotactile Perception with Active Vibration for Screw-Actuated Robot in Granular Media

Shoujie Li*, Changqing Guo*, *Student Member, IEEE*, Junhao Gong, *Student Member, IEEE*, Chenxin Liang, *Student Member, IEEE*, Wenhua Ding, Wenbo Ding, *Member, IEEE*

Abstract—Perception in granular media remains challenging due to unpredictable particle dynamics. To address this challenge, we present SandWorm, a biomimetic screw-actuated robot augmented by peristaltic motion to enhance locomotion, and SWTac, a novel event-based visuotactile sensor with an actively vibrated elastomer. The event camera is mechanically decoupled from vibrations by a spring isolation mechanism, enabling high-quality tactile imaging of both dynamic and stationary objects. For algorithm design, we propose an IMU-guided temporal filter to enhance imaging consistency, improving MSNR by 24%. Moreover, we systematically optimize SWTac with vibration parameters, event camera settings and elastomer properties. Motivated by asymmetric edge features, we also implement contact surface estimation by U-Net. Experimental validation demonstrates SWTac's 0.2 mm texture resolution, 98% stone classification accuracy, and 0.15 N force estimation error, while SandWorm demonstrates versatile locomotion (up to 12.5 mm/s) in challenging terrains, successfully executes pipeline dredging and subsurface exploration in complex granular media (observed 90% success rate). Field experiments further confirm the system's practical performance.

Index Terms—Visuotactile Perception, Event-based Vision, Granular Exploration, Screw-actuated Robot

I. INTRODUCTION

GRANULAR materials, such as sand, gravel, and powder, exhibit complex behaviors characterized by particle flowability, frictional interactions, and heterogeneous porosity [1]–[3]. These materials are commonly encountered in underground infrastructure inspection [4], mineral exploration [5], disaster rescue [6], and planetary missions [7]. In such applications, robotic systems often employ active techniques including drilling and vibration to facilitate penetration [8]–[11]. Due to the low visibility and unpredictable nature of these environments, tactile feedback is critical. It provides real-time information on contact forces and material properties, thereby enabling effective perception and adaptive control [12], [13].

Conventional electronic tactile sensors (e.g., strain gauges, piezoresistive arrays) would suffer performance degradation
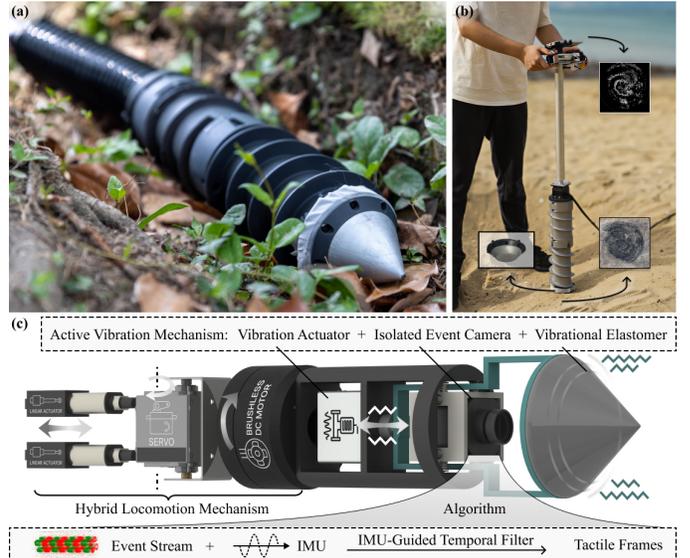
Fig. 1. Overview of SandWorm. (a) The snakelike robot is navigating an outdoor trench with the conical elastomer. (b) Manually operated drilling with the frustoconical elastomer on a sandy beach, showing the texture of an ammonoid fossil. (c) Schematic of SandWorm, featuring an integrated active vibration mechanism and hybrid locomotion system. An IMU-guided temporal filter is introduced to enhance imaging quality.

due to strong vibrational resonance in the drilling process [14], [15]. Visuotactile sensors reduce these mechanical drawbacks by separating the electronics from their sensing skin [16]. However, their reliance on the camera introduces new limitations, as high-frequency vibration would induce motion blur, image trailing, and feature-tracking errors [17], [18]. With the event-based imaging mechanism, event cameras can capture environmental changes in extremely short time intervals, effectively avoiding the blurring issues associated with traditional cameras due to long exposure times [19]. Moreover, with higher dynamic range, event cameras can detect subtle changes, making them ideal for precise visuotactile perception in granular media under vibration. However, event cameras intrinsically respond only to dynamic changes. To achieve static imaging, an imaging medium should be employed to generate detectable events [20].

Inspired by benthic polychaetes such as *alitta virens*, which use sensory-rich heads and muscle-driven peristalsis combined with spiral locomotion to traverse granular sediments [21], [22], we present SandWorm, a biomimetic robotic system. SandWorm mimics the worm's locomotion to explore granular media and confined environments, and integrates SWTac, an actively vibrated event-based visuotactile sensor for tactile

perception. The main contributions are as follows:

- **Active Vibration for Static Tactile Imaging:** We introduce a vibration-isolation mechanism in the SWTac visuotactile sensor that generates controlled elastomer vibration, enabling static tactile perception by the mechanically isolated event camera. This approach not only achieves pixel-level tactile imaging at 1000 Hz by overcoming the inherent limitation of event cameras in capturing stationary objects, but also leverages vibration-induced granular fluidization to reduce subsurface penetration resistance.

- **IMU-Guided Perception Algorithm :** To enhance event stream consistency, we propose an IMU-guided temporal filter based on a vibration-imaging model, achieving 24% improvement in masked signal-to-noise ratio (MSNR). Furthermore, leveraging asymmetric edge features verified by finite element analysis, we employ a U-Net model for contact surface estimation, achieving a structural similarity index measure (SSIM) about 0.97. We also systematically characterize the effects of vibration frequency, amplitude, direction, event camera sensitivity, and elastomer viscoelasticity in SWTac.

- **Hybrid Locomotion Mechanism:** We introduce a hybrid locomotion mechanism that synergistically combines screw-driven rotation with pushrod-actuated peristalsis for enhanced propulsion in confined environments. Implemented in SandWorm, a snakelike robot integrated with SWTac, this mechanism improves propulsion efficiency while enabling adaptable application.

- **Real-World Field Validation:** Extensive experiments demonstrate SWTac's capabilities in precise force estimation (0.15 N accuracy), high-resolution texture sensing (0.2 mm resolution), and robust stone classification (98% accuracy). SandWorm demonstrates robust performance across diverse field scenarios, successfully executing high-curvature movements, steering and multi-terrain traversal (dense grass, tangled bushes, cement surfaces) at speeds up to 12.5 mm/s; performing complex tasks such as pipeline dredging and subsurface exploration in non-uniform media (beach sand, garden soil, four industrial granular media); and achieving an observed 90% success rate in detecting subsurface objects.

The remainder of this paper is organized as follows. Section II reviews related works. Section III describes the design and framework of SWTac and SandWorm. Section IV presents the algorithms for enhancing event-based imaging, and characterization of SWTac. Section V details the experimental evaluations. Section VI discusses insights and future work. Finally, Section VII concludes this article.

II. RELATED WORKS

A. Perception in Granular Media

Perception in granular media poses unique challenges that have motivated a variety of non-invasive and invasive solutions. Current non-invasive methods, such as LiDAR [39], ultrasonic wave propagation [40], and Rayleigh waves [41] struggle in granular environments due to factors like clutter, material heterogeneity, and deformation under stress. Techniques like MRI [42] and dynamic X-ray radiography [43], are useful for inferring media properties rather than detecting subsurface objects.

Invasive sensing is therefore essential for subsurface exploration. Table I summarizes existing invasive robotic perception approaches. Robots operating in granular media [23], [24] and confined environments [25], [26] typically rely on cameras for perception, but these systems can only capture surface features and fail to image once the camera is in contact with the medium due to blocked illumination. Specialized invasive sensors are designed with penetration ability, but may suffer from prohibitively long detection times [30]. Force sensor [28], [29] and IMU [27] are commonly used for locomotion control, yet they measure only localized contact information and cannot

TABLE I
COMPARISON OF INVASIVE ROBOTIC PERCEPTION METHODS

Reference	Sensor	Resolution	Force & Contact
Richter et al. [23]	Camera	Pixel Level, ∼30 Hz	✗
Xue et al. [24]	Camera	Pixel Level, ∼30 Hz	✗
Kwon et al. [25]	Camera	Pixel Level, ∼30 Hz	✗
Kakogawa et al. [26]	Camera	Pixel Level, ∼30 Hz	✗
Nourizadeh et al. [27]	IMU	Single Point, 5 Hz	✗
Kolvenbach et al. [28]	Force	Single Point, 2 Hz	✓
Ruppert et al. [29]	Force	Single Point, 330 Hz	✓
Russell et al. [30]	Chemistry	Single Point, 0.003 Hz	✗
Digger Finger [8]	Visuotactile	Pixel Level, ∼30 Hz	✓
SandWorm (Ours)	Visuotactile	**Pixel Level, 1000 Hz**	✓

TABLE II
COMPARISON OF EVENT-BASED SENSORS FOR ROBOTIC APPLICATIONS

| Sensor	Latency	Spatial Resolution	Force Sensing	Sensing Type	Camera	Texture	Characteristics
Noise2Image [31]	100 ms	N/A	N/A	Vision	EV (Dynamic + Static)	N/A	Static Imaging for EV, Software
Yousefzadeh et al. [32]	100 ms	1 mm	✗	Vision	EV (Dynamic + Static)	✓	Static Imaging for EV, Hardware
AMI-EV [20]	**0.8 ms**	**0.04 mm**	✗	Vision	EV (Dynamic + Static)	✓	Static Imaging for EV, Hardware
NeuTouch [33]	10 ms	1 mm	✓	Vision, Tactile	RGB + EV (Dynamic)	✓	Multimodal Processing
Baghaei et al. [34]	21 ms	0.2 mm	0.16 N	Vision, Tactile	EV (Dynamic)	✓	Force Estimation
GelEvent [35]	5.5 ms	225 markers (3 mm)	0.8 N	Visuotactile	EV (Dynamic)	✗	Contact Area Estimation
E-BTS [36]	2 ms	5 markers	Shear 0.27 N	Visuotactile	EV (Dynamic)	✗	Teleoperation
Evetac [37]	1 ms	63 markers	Shear 0.22 N	Visuotactile	EV (Dynamic)	✗	Slip Detection
Digger Finger [8]	33 ms	0.3 mm	✓	Visuotactile	RGB (Static)	✓	GelSight [38] in Granular Media
SWTac (**Ours**)	**1 ms**	**0.2 mm**	**Shear 0.15 N**	Visuotactile	EV (Dynamic + Static)	✓	Granular Media Static Imaging for EV, Hardware
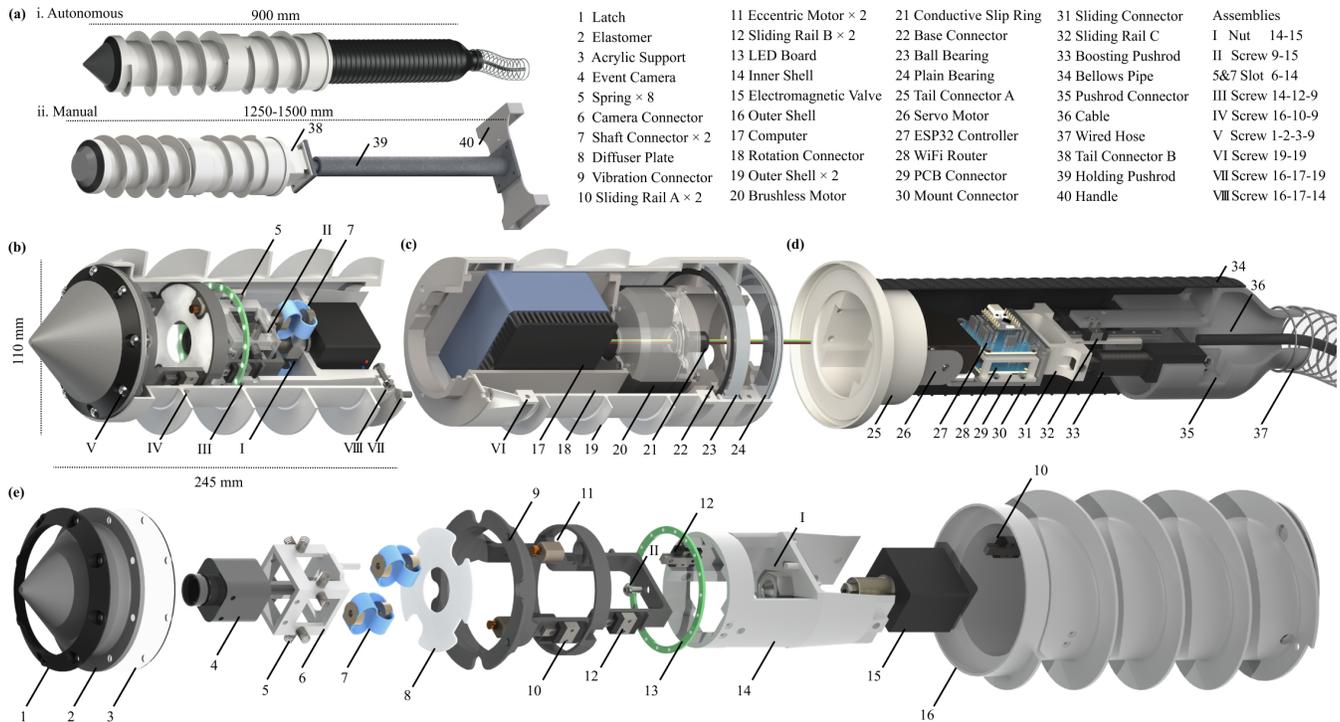
Fig. 2. Overview of SandWorm's mechanical structure. (a) Rendered views of SandWorm: an autonomous snakelike robot and a manually operated device. Cross-sectional views of (b) SWTac, the perception module, (c) the rotational driving module and (d) the snakelike tail module. (e) Exploded view of SWTac.

provide the broad coverage like vision-based methods. In contrast, visuotactile sensors integrate illumination, imaging, and contact skin into a single module [16], making them ideal for exploration in granular media. For example, DiggerFinger [8] performs exploration by GelSight [38] in a sandy environment, but its imaging performance would degrade under the strong vibrations required for medium penetration.

B. Visuotactile Sensors with Event Camera

Event cameras effectively mitigate motion blur, which is a common issue for traditional RGB cameras in dynamic environments [19], [44], such as those arising from vibration-driven exploration. As shown in Table II, existing event-based visuotactile sensors generally based on markers [35]–[37]. They follow the design of conventional visuotactile sensors (e.g. GelSight [38]), only replacing the RGB camera with event camera. By employing advanced elastomer designs, they achieve higher spatial resolution for tasks including force and contact surface estimation [35]. Although these sensors are particularly useful in teleoperation [36] and sliding detection [37], their reliance on markers obstructs the field of view and limits imaging to taxel-level resolution rather than pixel-level [8], resulting in poor texture perception and therefore unsuitable for exploration tasks.

If not relying on markers, existing methods fix the event camera externally to the tactile sensor to capture the dynamic deformations of the sensing skin during contact [33], [34]. Although effective for texture extraction, these approaches produce experimental setups to make the event camera remaining stationary, rather than proposing a compact, integrated visuotactile sensor. The primary challenge lies in the fact that event cameras only capture dynamic changes [19]. In integrated visuotactile systems, the sensing skin and event camera remain stationary relative to each other, so events are generated only during transient deformations [35], resulting in sparse and inconsistent imaging. Markers have predetermined positions, enabling localization of taxels despite these gaps. But in visuotactile sensors without markers, the absence of predictable features makes it difficult to acquire high-quality and consistent event-based tactile images.

Several vision-based solutions have been proposed to address the intrinsic limitation on static imaging of event cameras. Software-based noise sampling recovery methods [31] demand high processing time. For hardware-based solutions, directly vibrating the event camera carries the risk of mechanical damage [32]. A better approach is to add an imaging medium outside the sensor to trigger events, such as using a rotating prism to redirect the light path while the camera remains static [20]. Particularly, in visuotactile sensors, the elastomer could serve as a natural medium between the event camera and the object. Therefore, without requiring additional imaging medium, introducing active vibration to the elastomer would enable the event-based visuotactile sensor to capture a high-quality and consistent tactile event stream.

In this article, we present SandWorm, a robot designed for navigation and exploration in both granular media and confined environments (Fig. 2(a)). SandWorm comprises two main modules: an actuation module featuring a screw-actuated body enhanced with peristaltic motion for improved mobility, and a perception module, SWTac. SWTac is an active event-

based visuotactile sensor that applies controlled vibration to the elastomer while mechanically isolating the event camera (Fig. 2(b)). This active vibration effectively integrates perception and actuation, as it guarantees consistent tactile perception and simultaneously facilitates invasive exploration by loosening the granular material.

III. HARDWARE STRUCTURE

This section first details the hardware structure of SWTac, including the vibration-isolation mechanism, elastomer, and illumination design. We then introduce the actuation module, followed by an analysis of the rotational–peristaltic locomotion mechanics. Finally, we summarize SandWorm's perception-actuation framework to highlight our core innovative modules and their interconnections.

A. Perception Module Design

The structure of SWTac is designed to efficiently transmit active vibrations to the elastomer while simultaneously isolating the event camera. Fig. 2(b) shows the cross-sectional view of SWTac, and Fig. 2(e) presents a detailed schematic breakdown of its components. The core of this visuotactile sensor is the DVXplorer Mini event camera, selected for its compact size. The peripheral structure is designed around this camera and features two functionally distinct sections: a vibrational part and an isolated part.

For the vibrational part, the design aims to generate and guide multi-axis vibration using a compact assembly. This is achieved by mounting both the primary vertical actuator (electromagnetic valve, P15) and the supplementary horizontal actuators (eccentric motors, P11) onto a shared vibration connector (P9). This entire dynamic assembly is then coupled to the stationary inner shell (P14) via sliding rails (P10, P12). These rails function to constrain the motion primarily to the vertical axis while accommodating assembly tolerances, ensuring stable operation.

For the isolated part, the core design is to mechanically decouple the event camera (P4) from the high-vibration assembly. This is accomplished by soft-coupling the camera (mounted on P6) to the stationary inner shell using a dedicated multi-axis isolation system: eight springs (P5) provide horizontal isolation, and two flexible shaft connectors (P7) provide vertical isolation. This design effectively protects the sensitive electronics without obstructing the camera's field of view. Finally, the LED board (P13) and diffuser plate (P8) are integrated for illumination. The whole assembly is secured by the primary load-bearing (Assembly VIII) and enclosed within the spiral outer shell (P16) to enhance drilling performance.

The elastomer serves as a fundamental component in visuotactile sensors. In granular media, elevated abrasive forces necessitate a thicker elastomer layer than those used in conventional designs to enhance durability. Therefore, polydimethylsiloxane (PDMS; Sylgard™ 184, Dow Corning) is adopted to guarantee optical transparency. The assembled elastomer tip (P1-P3) is installed at the end of the vibration connector. The vibration-isolation structure enables dual-axis vibrational control of the elastomer tip, both horizontally and vertically.
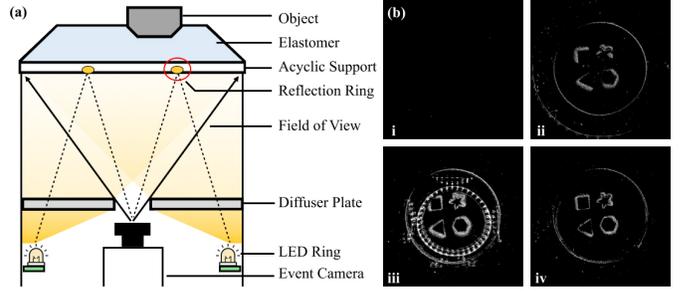

Fig. 3. Illumination setup and conditions. (a) Hardware setup and optical-path layout. (b) Reconstructed images of event camera with i. No illumination; ii. Direct dark-field illumination; iii. Direct bright-field illumination; iv. Diffuse bright-field illumination with a diffuser plate.

The illumination system is crucial to event cameras, as sufficient lighting is essential for enhancing imaging quality. In our design, we utilize a circular array of 14 LEDs (WS2812B 5V) with a diffuser plate. As illustrated in Fig. 3, with direct lighting, a peripheral reflective ring is observed. Direct dark-field illumination would reduce the quality of the primary signal by diverting the light path from a direct frontal configuration to a lateral one. Fig. 3(b)iv shows the capture result by diffuse bright-field illumination. The diffuser plate homogenizes the light distribution while attenuating its intensity, thereby reducing the severity of specular reflections. Additionally, the central opening of the diffuser plate is also optimized to provide a hardware feature mask that delineates the region of interest for subsequent software processing.

B. Actuation Module Design

Snakelike robots are commonly used for exploration in granular environments. These robots utilize different actuating methods, including lateral wave motion [45], peristaltic motion [46], [47] and screw mechanisms [7], [23]. Screw mechanisms provide both propulsion and steering forces, making them particularly effective for navigating in confined spaces [48] and pipelines [23]. Therefore, we adopt a rotational-peristaltic driving scheme for locomotion, which requires fewer actuators compared to lateral wave methods and enhances traction in granular media. As depicted in Fig. 2(a), SandWorm has a reconfigurable design, and can work either as a autonomous snakelike robot or as a manually operated device. Both configurations differ only at the tail end, while the perception module (SWTac, Fig. 2(b)) and the rotational driving module (Fig. 2(c)) are identical.

The rotational driving module (Fig. 2(c)) provides both data processing and screw actuation. An onboard computer (P17) executes event–stream processing and high-level controlling. A brushless motor (P20) is employed to continuously rotating the SWTac sensor. The drivetrain incorporates a coaxial bearing stack and a dedicated mechanical interface, ensuring clean separation between the rotating and non-rotating modules. A spiral outer shell surrounds the assembly, implemented as two semi-circular housings (P19, Assembly VI) that enclose the periphery of the module. The shell is machined from an Al–Mg alloy to improve structural robustness and wear resistance during contact with granular media.

Fig. 4. Overview of the SandWorm robot's DOFs. Locomotion is achieved by combining rotation, peristaltic translation, and vertical oscillation. The vibrational elastomer tip for perception is also illustrated.
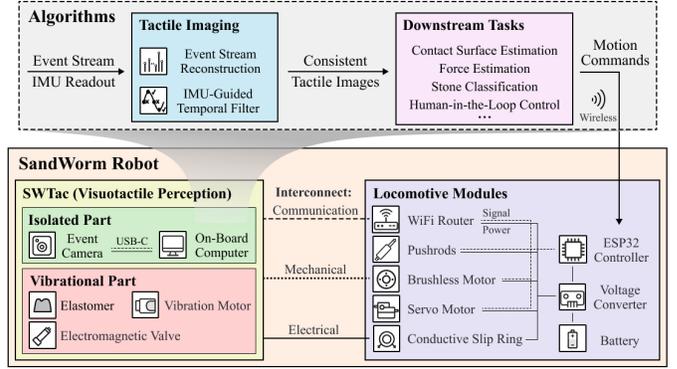

Fig. 5. System architecture of the SandWorm robot, detailing the hardware and algorithm integration. The hardware (bottom) features the SWTac visuotactile sensor (comprising an isolated and a vibrational part) and the locomotive modules. The algorithmic pipeline (top) processes the event stream and IMU readout from the event camera, including an IMU-guided temporal filter followed by downstream tasks.

The tail module (Fig. 2(d)) hosts the remaining actuation. It embeds electronics within a collapsible bellows conduit (P34). Steering actuation is provided by a servo motor (P26), while peristaltic motion is generated by a pushrod mechanism (P33). An ESP32 microcontroller (P27) and a compact router (P28) serve as the low-level controller and wireless communication backbone. Linear guides (P32) and reinforced metal connector (P30) increase stiffness under load. At the distal end, a compliant, non-extendable support hose (P37) with cable (P36) inside is employed. For extended applications, a pushrod (P39) of larger size will be mounted to the rotational driving module, forming a hand-held (P40) drilling assembly.

C. Locomotion Mechanics

SandWorm provides three degrees of freedom (DOFs) of locomotion, as shown in Fig. 4: a rotation of the spiral shell driven by a brushless motor; a servo-driven vertical oscillation for steering; and a peristalsis motion via internal pushrods for linear translation. The front rotational axis is rigidly coupled to SWTac enclosed by the spiral shell. The rear of SandWorm is guided by a flexible hose which supplies reaction force. Consequently, with its tail serving as a pivot, SandWorm leverages frictional contact to realize forward motion.

To realize robust locomotion under frictional contact, Sand-Worm employs a peristaltic drive made of two alternating phases that combine screw traction with pushrod actuation. In confined environments (e.g., pipelines), gravity together with the pipe support provides roll stability so that a single-chirality screw can generate forward thrust without requiring paired counter-rotating screws. Let F_p denote the axial force applied by the pushrod and F_friction the lumped friction opposing motion. The effective gravitationally resolved driving component along the axis is denoted F_propel (cf. Appendix A). During the extension phase, the pushrod assists propulsion, leading to the net force:

$$F_\mathrm{extension} = F_\mathrm{propel} + F_\mathrm{p} - F_\mathrm{friction}, \tag{1}$$

which produces a rapid forward stroke. During retraction, the pushrod opposes motion, and the net force becomes:

$$F_\mathrm{retraction} = F_\mathrm{propel} - F_\mathrm{p} - F_\mathrm{friction}, \tag{2}$$

yielding little displacement. This asymmetric two-phase actuation, akin to a controlled shift of the robot's center of mass,

produces a net advance greater than that of continuous rotation alone.

D. Perception-Actuation Framework

The system architecture of SandWorm, as depicted in Fig. 5, integrates the SWTac visuotactile sensor with locomotive modules, as well as a hierarchical signal processing pipeline. The perception component is the SWTac sensor. Its vibrational part includes the actuators, while its isolated part contains the event camera and the on-board computer. The robot's actuation is handled by the locomotive modules, comprising pushrods, a brushless motor, and a servo motor, all managed by an ESP32 controller and a central battery. The interconnection between the SWTac and the locomotive modules is threefold. Brushless motor mechanically couples the sensor to the locomotive body. Conductive slip ring delivers power from the non-rotating locomotive section to the rotating SWTac as the electrical link. The communication link connects the on-board computer to the ESP32 by WiFi.

The algorithmic pipeline (top panel) runs entirely on the SWTac's on-board computer. Raw sensor data, specifically the event stream and IMU readout, are used to reconstruct frames and are processed by our IMU-guided temporal filter, which generates consistent tactile images. This processed data is then passed to the downstream tasks (e.g., contact surface estimation, force estimation) to generate high-level motion commands, either directly or indirectly through human-in-the-loop control (utilizing remote real-time image viewing). These commands are transmitted wirelessly via the communication interconnect to the ESP32 controller in the locomotive modules, which in turn actuate the motors and pushrods. Thus, the SWTac's tactile perception, which relies on vibration and motion, acts back on the actuators to form a fundamental perception-actuation loop.

IV. ALGORITHM DESIGN

In this section, we first introduce the workflow to reconstruct grayscale images from the asynchronous event stream.
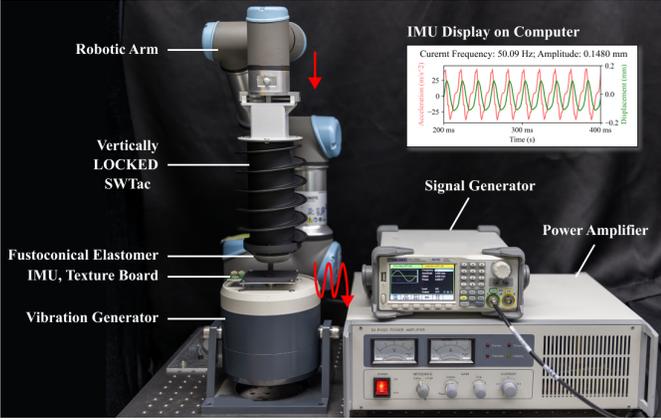
Fig. 6. Data acquisition platform. SWTac with frustoconical elastomer is connected to the robotic arm. The texture board and IMU are attached to a programmable vibration generator.
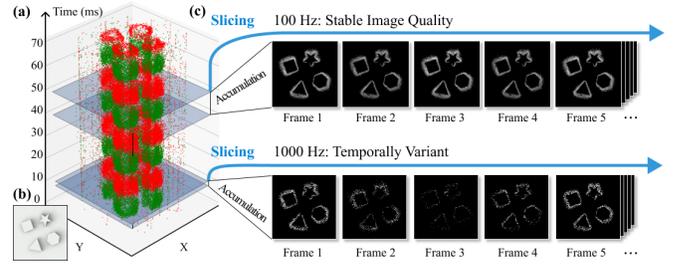

Fig. 7. Principle and results of grayscale image reconstruction. (a) Event stream during vibration, with positive events in red and negative events in green. (b) Texture board. (c) Grayscale images under different reconstruction frame rates.

To mitigate vibration-induced temporal inconsistencies, we develop an IMU-guided temporal filter based on the theoretical model of event-based imaging under vibration. After that, motivated by the asymmetric edge features in reconstructed tactile images, we conduct contact surface estimation, beyond the edge information captured by the event camera. Finally, we characterize the imaging performance of SWTac by different parameters, including vibration amplitude, frequency, direction, event camera sensitivity and elastomer viscoelasticity.

A. Event-based Image Reconstruction

1) Hardware Setup: To enhance perception performance by algorithm design, we set up a data acquisition platform in Fig. 6. Instead of using the electromagnetic valve, we utilized a vibration generator to provide active vibration. This generator is driven by a signal generator with a power amplifier, providing explicit frequency control and implicit amplitude control via adjustable power of stable sine wave. An IMU mounted on the generator communicates with a computer in real time to compute the amplitude, enabling adjustments. A UR5 robotic arm was employed to position the sensor accurately above the vibration generator.

To ensure the effect of the external vibration generator is equivalent to the sensor's active vibration, we modified the sensor housing by removing the vertical elastic DOF, thereby rigidly securing the elastomer. Externally induced vibrations technically would produce the same effect as actively vibrating the elastomer, given that the sensor remains stationary in both setups. We verified the interchangeability of these two actuators in Section V-A. The frustoconical elastomer with a flat surface was used for data collection.

For the event camera settings, we used three threshold levels of the camera change-detection circuitry to assess sensitivity [49]. The threshold defines the minimal log intensity variation at the pixel level required to induce an event. Lower thresholds enhance sensitivity by generating more events in response to subtle light changes, while higher thresholds diminish sensitivity by excluding minor intensity variations.

We employed these three threshold values to capture event streams corresponding to different camera sensitivity settings.

2) Reconstruction Workflow: The captured event streams are sequences of discrete events $\mathcal{E} = \{e_i\}$, each encoding (x_i, y_i, p_i, t_i) as a pixel coordinate, a brightness-change polarity and a precise timestamp. An example event stream is depicted in Fig. 7(a). Under the active vibration of 50 Hz in SWTac, the temporal sequence of the event stream shows an alternating pattern with color-encoded polarity.

Next, the event stream is processed to a sequence of frame-based grayscale images G_k, as shown in Fig. 7(c). This conversion is achieved by event accumulation. As a pre-processing step, an active background filter was adopted to suppress noise from the raw stream, yielding a filtered event set $\mathcal{E}_{\text{filtered}}$. Subsequently, we convert these asynchronous events into frames at f_{rate} frame rate by slicing the filtered stream into discrete time windows of a specific accumulation time, $\Delta T_{\text{acc}} = 1/f_{rate}$, and projecting the events within each slice onto the X-Y plane. Each event contributes a fixed numeric value C, namely the contribution, to its pixel's brightness. This process is formally defined as:

$$G_k(x,y) = \sum_{\substack{e_i \in \mathcal{E}_{\text{filtered}} \\ (x_i, y_i) = (x,y) \\ t_i \in [T_k, T_k + \Delta T_{\text{acc}})}} C \qquad (3)$$

where $G_k(x,y)$ is the intensity of frame k at pixel (x,y), and T_k is the start time of the frame. This formulation inherently applies a step-function reset between windows, preventing carry-over artifacts. Moreover, by discarding polarity information (the p_i term is omitted), we rely solely on event counts per pixel, which further sharpens edges in the reconstructed frames. The data processing of event camera was primarily conducted using the DV-processing library [50].

TABLE III
PARAMETER SETTINGS FOR GRAYSCALE IMAGE RECONSTRUCTION

Accumulation Time	Contribution	Reconstruction Rate
1 ms	1	1000 Hz
2 ms	1	500 Hz
5 ms	0.5	200 Hz
10 ms	0.25	100 Hz
33 ms	0.08	30 Hz

With carefully chosen parameters for accumulation time and contribution, we could achieve real-time grayscale reconstruction at up to 1000 Hz, as listed in Table III. However, the alternating pattern of the event stream directly influences the reconstruction quality. Fig. 7(c) shows that a low reconstruction rate (e.g., 100 Hz) produces consistent images that do not capture the vibration pattern. In contrast, a higher reconstruction rate (e.g., 1000 Hz) captures the dynamic variations of the vibration pattern, resulting in image inconsistency.

3) Imaging Quality Metrics: To assess the quality of the reconstructed grayscale images, we introduce MSNR, Shannon entropy and MSE. MSNR is defined as the masked signal-to-noise ratio calculated by excluding most of the uninformative black background. Specifically, we compute the SNR only within the foreground region Ω as follows:

$$\text{MSNR} = 10 \lg \left(\frac{\sum_{(i) \in \Omega} I(i)^2}{\sum_{(i) \in \Omega} |I(i) - \mu_\Omega|^2} \times \frac{N_\Omega}{N_{\text{Image}}} \right), \quad (4)$$

where $I(i)$ is the grayscale value of the i-th pixel in the image, and μ_Ω is the mean grayscale value over Ω. N_Ω and N_{Image} represent the number of pixels in region Ω and the image.

The Shannon entropy S is adopted to quantify the information content of the image:

$$S = -\sum_{l=0}^{L-1} q_l \log_2 q_l, \quad (5)$$

where q_l is the occurrence probability of the l-th grayscale level and L is the number of grayscale levels.

The MSE metric is used to evaluate the similarity between the reconstructed image and the ground truth edges. To establish the groundtruth image, we first extract the edges from the corresponding texture board model, then compute the undistorted coordinate transformation by camera calibration. The MSE is calculated as:

$$\text{MSE} = \frac{1}{N_\Omega} \sum_{i=1}^{N_\Omega} |I(i) - I_{\text{GT}}(i)|^2, \quad (6)$$

where $I_{\text{GT}}(i)$ is the corresponding pixel value of ground truth.

B. IMU-Guided Temporal Filter

In this part, we propose the IMU-guided temporal filter, addressing the temporal inconsistency of the event stream. We formulated and validated the mathematical model of event-based visuotactile perception with active vibration, proposed a deployment strategy and analyzed the statistical results.

1) Motivation: While an active vibration strategy is essential to make static objects visible to the event camera, it also introduces oscillatory interaction between the elastomer and external objects. Since event cameras are highly sensitive to these temporal variations, the vibration induces fluctuations in event density and imaging quality when using a high reconstruction rate. As seen in Fig. 7(c), this results in a temporally variant stream where high-quality frames are interleaved with sparse, low-quality frames, degrading reconstruction stability.

Therefore, it's necessary to automatically retain only the high-quality segments of the event stream. To achieve this,

we leverage the real-time acceleration data from the event camera's internal IMU as a prior. Because the acceleration is rigidly coupled to the vibration mechanism, it could provide a precise estimation of the elastomer's current vibration phase and, by extension, the expected imaging quality in real-time. We use this information to develop a model-based algorithm that guides a temporal gate: the filter actively retains the event stream during high-quality phases and discards the stream during low-quality phases. This process ensures the final reconstructed images are built only from high-quality data with enhanced consistency.

2) Mathematical Modeling: To establish a model for event-based tactile imaging under active vibration, we start by analyzing the relationship between vibrational contact and the quality of reconstructed images. For simplification, we analyze the vibration pattern using only vertical vibration. The vibration excitation model of the elastomer can be expressed as a sinusoidal function. During tactile interactions between the elastomer and external objects, several factors including material properties, deformation characteristics, and system hysteresis would cause variations in amplitude and introduce a phase delay. The resulting vibration model can be modelled as:

$$f(t) = A \sin(\omega t + b), \quad (7)$$

where f, A, ω, t, and b are the displacement, amplitude, the angular frequency, time, and the system's phase delay, respectively.

In the imaging model of event cameras, we use an abstract metric, imaging quality (IQ), to analyze the performance. Specific metrics for evaluation will be provided in the next section. The characteristic of event cameras is to capture changes in illumination, and therefore, IQ tends to increase with the absolute vibration velocity of the imaging medium, i.e., the elastomer. Therefore, we assume a proportional relationship for simplicity in modeling as:

$$\text{IQ}(t) \approx u \left| \frac{\mathrm{d}f(t)}{\mathrm{d}t} \right| = |uA\omega \cos(\omega t + b)|. \quad (8)$$

where u is a scaling factor. By trigonometric identities and Fourier series, the above equation can be rewritten in a standard sinusoidal form:

$$\text{IQ}(t) = \frac{2|uA\omega|}{\pi} + \frac{4|uA\omega|}{3\pi} \sin(2\omega t + 2b + \frac{\pi}{2}), \quad (9)$$

which indicates that the imaging quality signal contains a sinusoidal component, a phase delay, and a constant offset.

In practice, the velocity is difficult to measure directly, and acceleration data from the IMU are typically used instead. The IMU measurement corresponds to the second derivative of displacement:

$$\text{IMU}(t) = \frac{\mathrm{d}^2 f(t)}{\mathrm{d}t^2} + \epsilon(t) = A\omega^2 \sin(\omega t + \pi) + \epsilon(t). \quad (10)$$

where the $\epsilon(t)$ term, representing the noise and non-periodic elements, which could be effectively handled by band-pass filtering.

At this point, we can make an inference between the vibration and imaging models. Comparing the IQ and IMU signals, two important relationships are utilized:
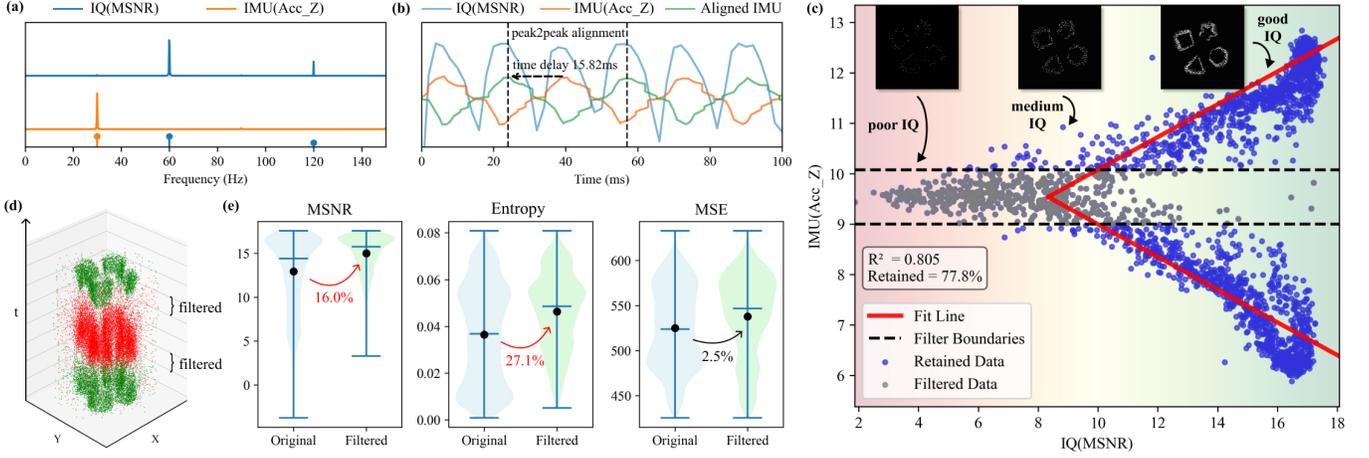
Fig. 8. Validation of the IMU-guided temporal filter. (a) Fourier transform analysis of the imaging quality metric alongside IMU signal. (b) Peak-to-peak alignment. (c) Relations between IMU and MSNR, illustrating the regression result, filter boundaries and representing frames. (d) Intuitive illustration of IMU-guided temporal filtering on the event stream. (e) Statistical comparison of imaging quality metrics before and after filtering.

- The signal frequency of IQ is explicitly twice that of the IMU signal.
- There is a stable but implicitly phase difference between the two signals.

Therefore, a temporal alignment between the two signals is required to estimate the optimal time shift $\hat{\Delta}t$. The theoretical relationship between the filtered IMU signal and the imaging quality can thus be approximated as:

$$\text{IQ}(t) = k \cdot \left| \tilde{\text{IMU}}(t - \hat{\Delta}t) \right| + c, \tag{11}$$

where k is a scaling factor and c is a constant offset, determined experimentally. Here, $\tilde{\text{IMU}}(\cdot)$ represents the IMU signal component obtained after applying a band-pass filter centered at the active vibration frequency ω.

3) Verification: Based on the theoretical analysis, we performed frequency validation, phase alignment, regression fitting to implement the calibration process of the IMU-guided temporal filter. We used data acquired under a 30 Hz vertical vibration with a reconstruction rate of 500 Hz. MSNR is used as the representing imaging quality metric and computed for each image, forming a time series $\text{IQ}(t)$ of the total 2500 grayscale images in 5 s. In parallel, the IMU data, along with their corresponding timestamps, were extracted from the raw event camera output. We use the Z-axis of the IMU data, which represents the vertical vibration, denoted as Acc_Z.

First, we applied Fourier transform analysis to both the MSNR series and the IMU measurements. Fig. 8(a) shows that the dominant frequency in the imaging quality metric is 60 Hz, precisely twice of the 30 Hz vibration frequency in the IMU signal, which conforms to our theoretical predictions. Next, we aligned the timestamps of the MSNR series and the IMU signal by a peak-to-peak matching method. We first detecting the peak locations of both IMU and imaging quality signals, and then pairing each IMU peak with its nearest imaging quality peak to determine the average temporal delay $\hat{\Delta}t$:

$$\hat{\Delta}t = \frac{1}{N} \sum_{n=1}^{N} \left(t^{(n)}_{\text{IQ peak matched}} - t^{(n)}_{\text{IMU peak}} \right), \tag{12}$$

As depicted in Fig. 8(b), the 100 ms window shows the frequency doubling and time shift characteristics, verifying the temporal relationship.

Once the signals are temporally aligned, we pair each reconstructed frame's MSNR metric with the temporally matched IMU measurement. The images with higher MSNR visually have more complete and distinct edges. Specifically, we calculate the mean of the IMU measurements in the Z-axis over the corresponding interval of accumulation. As shown in Fig. 8(c), the scatter plot of MSNR versus the corresponding IMU measurements displays a clear correlation. A regression analysis was performed based on Eqn. (11), yielding an R^2 value of 0.805. This result validates our theoretical analysis and demonstrates that the IMU measurements can reliably predict the imaging quality of reconstructed frames.

4) Deployment Strategy: For the purpose of filtering, a minimum acceptable imaging quality threshold ($\text{IQ}_{\text{thresh}}$) is defined. In this experiment, for instance, the average MSNR measurement of 10.00 is adopted. Based on the regression analysis of the calibration data, this $\text{IQ}_{\text{thresh}}$ is mapped to a corresponding set of IMU measurement thresholds. Specifically, it's more than 10.08 or less than 9.00 in this example. Using this criterion, poor-quality segments of the event stream can be identified and filtered out, as intuitively illustrated in Fig. 8(d). The result in Fig. 8(c) highlights the retained frames in blue, where 77.8% of the data were preserved.

The procedure of the IMU-guided temporal filter is summarized in Algorithm 1. Rather than directly accumulating events, the IMU measurements embedded in the event camera were used as guidance to filter out the poor streams. It operates in two distinct phases: a one-time calibration and a real-time application. For calibration, it determines the optimal temporal alignment $\hat{\Delta}t$ (Eqn. (12)) and the IQ (Eqn. (4)) threshold $\text{IQ}_{\text{thresh}}$ based on the reconstructed frames (Eqn. (3)). Through regression analysis (Eqn. (11)), this quality threshold is then mapped to a corresponding motion magnitude threshold, $\tilde{\text{IMU}}_{\text{thresh}}$. Following calibration, the main filter iterates through the remaining event stream. It interpolates

the filtered and aligned IMU value ($\widetilde{\text{IMU}}_{\text{live}}$) and compares its magnitude to $\widetilde{\text{IMU}}_{\text{thresh}}$. Only segments that satisfy the criterion ($|\widetilde{\text{IMU}}_{\text{live}}| \geq \widetilde{\text{IMU}}_{\text{thresh}}$) are reconstructed (Eqn. (3)) and outputted, effectively discarding segments predicted to have poor imaging quality.

Algorithm 1 IMU-Guided Temporal Filter.

CALIBRATE($\mathcal{E}_{\text{cal}}$, IMU$_{\text{cal}}$, f_{rate}, ω)
    (G_{cal}, T_{frames}) $\leftarrow$ Reconstruction($\mathcal{E}_{\text{cal}}$, f_{rate}) by Eqn. (3)
    IQ$_{\text{series}}$ $\leftarrow$ Calculate_IQ(G_{cal}) by Eqn. (4)
    IQ$_{\text{thresh}}$ $\leftarrow$ Mean(IQ$_{\text{series}}$)
    $\widetilde{\text{IMU}}_{\text{signal}}$ $\leftarrow$ Bandpass(IMU$_{\text{cal}}$, ω)
    $\widetilde{\text{IMU}}_{\text{series}}$ $\leftarrow$ Sample($\widetilde{\text{IMU}}_{\text{signal}}$, T_{frames})
    $\hat{\Delta}t$ $\leftarrow$ Alignment(IQ$_{\text{series}}$, $\widetilde{\text{IMU}}_{\text{series}}$, T_{frames}) by Eqn. (12)
    (k, c) $\leftarrow$ Regression(IQ$_{\text{series}}$, $\widetilde{\text{IMU}}_{\text{series}}$, $\hat{\Delta}t$) by Eqn. (11)
    $\widetilde{\text{IMU}}_{\text{thresh}}$ $\leftarrow$ (IQ$_{\text{thresh}}$ $-$ c)/k
    return $\hat{\Delta}t$, $\widetilde{\text{IMU}}_{\text{thresh}}$

IMUGUIDEDFILTER($\mathcal{E}_{\text{stream}}$, IMU$_{\text{stream}}$, f_{rate}, ω)
    ($\mathcal{E}_{\text{cal}}$, IMU$_{\text{cal}}$) $\leftarrow$ GetFirstSegment($\mathcal{E}_{\text{stream}}$, IMU$_{\text{stream}}$)
    ($\hat{\Delta}t$, $\widetilde{\text{IMU}}_{\text{thresh}}$) $\leftarrow$ CALIBRATE($\mathcal{E}_{\text{cal}}$, IMU$_{\text{cal}}$, f_{rate}, ω)
    while $\mathcal{E}_{\text{stream}}$ has remaining segments
        t_{live} $\leftarrow$ Get_Timestamp($\mathcal{E}_{\text{stream}}$)
        $\widetilde{\text{IMU}}_{\text{live}}$ $\leftarrow$ Sample&Bandpass(IMU$_{\text{stream}}$, $t_{\text{live}} - \hat{\Delta}t$)
        if $|\widetilde{\text{IMU}}_{\text{live}}| \geq \widetilde{\text{IMU}}_{\text{thresh}}$
            G_{live} $\leftarrow$ Reconstruction($\mathcal{E}_{\text{stream}}$, f_{rate}) by Eqn. (3)
            Output G_{live}
        end if
    end while

Statistical analysis comparing the pre-filtered and post-filtered results (Fig. 8(e)) shows that the IMU-guided temporal filter substantially improves the MSNR and entropy, while MSE has a slight increment. Specifically for MSNR, an increase in the average value (Avg.) corresponds to an overall enhancement of imaging quality, while a reduction in the standard deviation (Std.) reflects improved stability and temporal consistency. Further experiments across various vibration frequencies and reconstruction rates (Table IV) demonstrate a maximum average increase of 24.0% and a standard deviation decrease of 45.8% in MSNR.

TABLE IV
MSNR ENHANCEMENT BY IMU-GUIDED TEMPORAL FILTER

Vibration Frequency	Reconstruction Rate	Processing Delay	Retaining Rate	Avg. Increase	Std. Decrease
30 Hz	500 Hz	2 ms	77.8%	16.0%	45.8%
30 Hz	1000 Hz	1 ms	81.1%	24.0%	40.8%
50 Hz	1000 Hz	1 ms	93.0%	11.6%	23.0%

C. Contact Surface Estimation

In Section IV-A, we noted that the event camera only highlights the edges of contacting objects. As a result, the reconstructed image underestimates the interior contact region and fragments the physical footprint. Therefore, it's necessary to estimate the full contact surface from the tactile image [33],
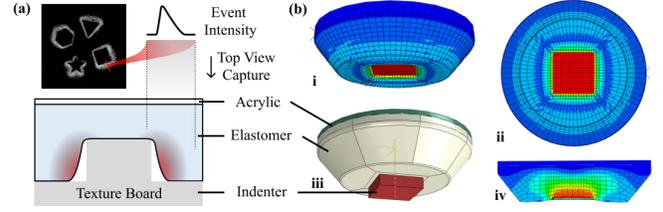

Fig. 9. Motivation and method of contact surface estimation. (a) Elastomer deformation under compression captured as asymmetric edge features. (b) Finite element analysis of elastomer deformation. iii. Simulation setup. Deformation contour plot in i. isometric; ii. top; iv. cross-sectional view of the elastomer.

[51], [52]. The capture edges arise from a relative height difference on the texture board. Due to the elastomer's continuity and elasticity, it cannot conform exactly to the vertical face of the texture, as illustrated in Fig. 9(a). Instead, the deformation is governed by Young's modulus and Poisson's ratio [53], [54]. It radiates outward from the contact edge until the stress relaxes. In the protrusion scenario, the deformation of elastomer creating a shadow that extends beyond the true boundary, while the contact side remains sharp. The event intensity curve clearly illustrates the asymmetric feature, with one side sharpened while the other blurred.

To further validate this asymmetric deformation, we performed finite element analysis (FEA) simulations using Abaqus. The model depicted in Fig. 9(b)iii featured a rectangular rigid indenter pressing onto the elastomer surface with 2 mm vertical displacement, employing surface-to-surface contact conditions. A fixed rigid acrylic plate fully constrains the elastomer from above. Elastomer parameters were set to a Young's modulus of 0.2 MPa and a Poisson's ratio of 0.48, using 1.5 mm C3D8R elements for meshing. The resulting deformation contour plot in Fig. 9(b) confirms the asymmetrical edge deformation. Stress rapidly accumulates near the inner contact boundary, creating an abrupt deformation edge (sharp imaging), while it gradually dissipates outward, resulting in a smooth deformation gradient (blurred imaging). The close agreement between simulation results and experimental imaging validates our assumption about asymmetric deformation patterns. Therefore, without needing markers, we can leverage this feature to recover the full geometry of the contact surface rather than merely estimating its area [35], [55].

D. Sensor Characterization

To characterize the SWTac sensor, we performed a statistical analysis on grayscale tactile images reconstructed from event streams. We systematically evaluated SWTac's performance by varying key parameters, including: vibration amplitude (0–400 μm), frequency (0–400 Hz), event sensitivity (three thresholds), vibration direction (vertical and horizontal) and elastomer property (viscoelasticity). A constant reconstruction rate of 100 Hz was used throughout the characterization.

1) Vibration Amplitude: First, we evaluated the sensor's performance variation with amplitude at three different frequencies, as shown in Fig. 10(a-c). The results indicate that when the amplitude is below 100 μm, the MSNR values
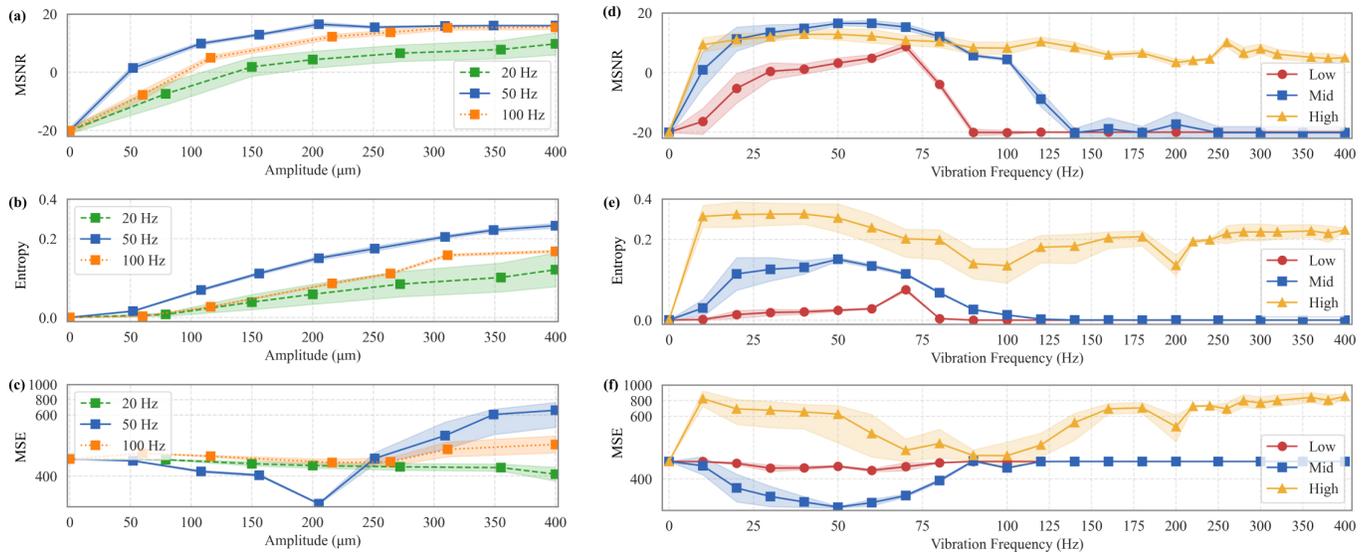
Fig. 10. Statistical evaluation of grayscale imaging quality under different vibrational conditions. (a–c) Amplitude variations with a mid sensitivity setting and three different frequencies. (d–f) Frequency variations with a 200 μm amplitude and three different event sensitivity settings.

are low but increase rapidly. Above 200 μm, the MSNR remains stable, but its variance increases. Notably, a medium vibration frequency of 50 Hz consistently provides superior performance.

The rationale for this behaviour is twofold. For lower amplitudes, the induced deformation is too small to generate sufficient information. While higher amplitudes initially capture more information (increasing entropy), amplitudes up to 200 μm yield signals that more closely approximate the ground truth (indicated by reduced MSE). Further increases in amplitude cause the signal to deviate due to higher reaction forces and noise. This is because the larger amplitude exceeds the elastomer's effective deformation range. Consequently, the apparent Young's modulus increases sharply, leading to smaller strain changes despite larger deformations. In short, although larger amplitudes capture more information, the added noise simultaneously reduces imaging quality.

2) Vibration Frequency: Next, we evaluated the sensor's performance under various vibration frequencies (0–400 Hz), as shown in Fig. 10(d-f). For frequencies between 0–100 Hz, the amplitude was kept at a constant 200 μm. While for frequencies between 100 and 400 Hz, the power amplification factor was fixed due to constraints of the IMU range. In the low-frequency range (0–100 Hz), the MSNR first gradually improves with increasing frequency, peaking at about 50 Hz. After this peak, the MSNR declines. At higher frequencies (100–400 Hz), damping effects become significant. For mid and low sensitivity, entropy drops to nearly zero, indicating no effective information is captured, and MSE approaches levels seen without vibration (0 Hz). For high sensitivity, MSNR remains relatively stable but is still worse than in the low-frequency range.

The initial performance increase up to 50 Hz is because the masked ground truth region accumulates more events, resulting in clearer edge signals (evidenced by higher entropy and lower MSE). The subsequent decline occurs because the

elastomer's damping effect begins to dominate over the active vibration. At higher frequencies, these damping effects and limited elastomer deformation cannot be fully compensated by post-processing, making high-frequency vibration unsuitable. Specifically, in the 20–40 Hz range, the higher standard deviation compared to 50 Hz is observable. This is attributed to the variance in reconstructed image, which arises from capturing the elastomer in different vibration phases. This is precisely the issue addressed in Section IV-B.

3) Event Camera Sensitivity: The selection of the event threshold presents a clear trade-off in Fig. 10(d-f). Our results show that the highest MSNR is achieved at approximately 50 Hz with a mid sensitivity setting. In the 20–80 Hz frequency range, the mid sensitivity consistently outperforms high sensitivity. At higher frequencies, although a higher sensitivity can capture texture patterns, the imaging quality still does not match the performance observed in the low-frequency range.

This trade-off is explained by the other metrics: a higher event threshold captures more information, reflected by higher entropy. However, it also produces thicker, more pronounced edges and increased background noise, leading to a higher MSE. In contrast, lower sensitivity reduces noise but may result in incomplete edge detection. The mid sensitivity setting, therefore, provides the optimal balance between information capture and noise induction.

4) Vibration Direction: We evaluated the effect of superimposing horizontal vibration (0, 50, 75, 100 Hz) onto vertical vibration (0, 20, 50, 100 Hz). The results in Fig. 11(a-c) show that introducing horizontal vibration generally enhances the MSNR, increases entropy, and reduces MSE. This improvement is especially significant when vertical vibration is less effective (e.g., at 0 Hz). An intuitive illustration is provided in Fig. 11(d).

The improvement is attributed to the fact that horizontal vibration leads to more uniform contact, enhancing edge integrity in the reconstruction images. Although the amplitude
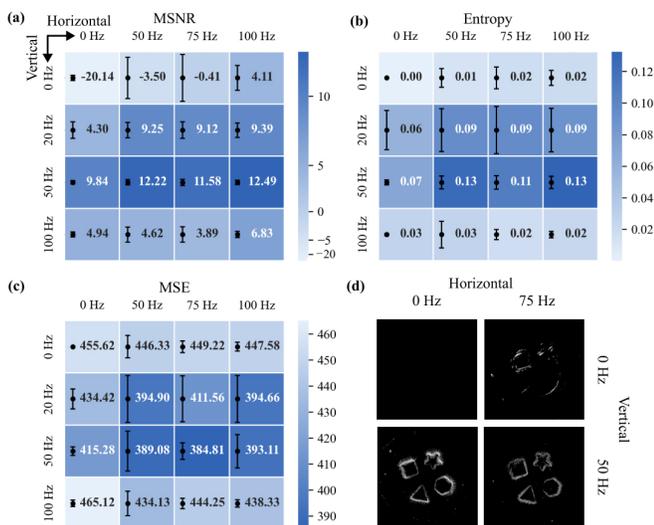
Fig. 11. Grayscale imaging quality under different vibration directions. (a–c) Imaging quality metrics under different frequencies. (d) Illustration of reconstructed grayscale images.
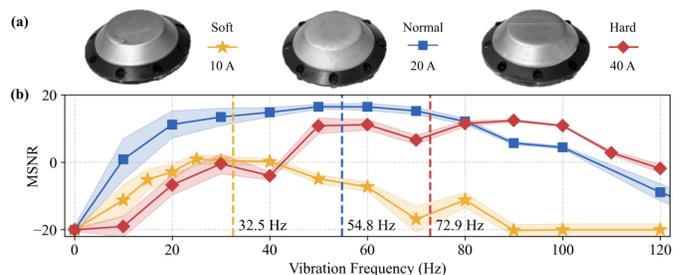

Fig. 12. Grayscale imaging quality for elastomers of different Shore hardness, representing different viscoelasticity. (a) Photographs of elastomers. (b) Imaging quality (e.g. MSNR) versus vibration frequency.
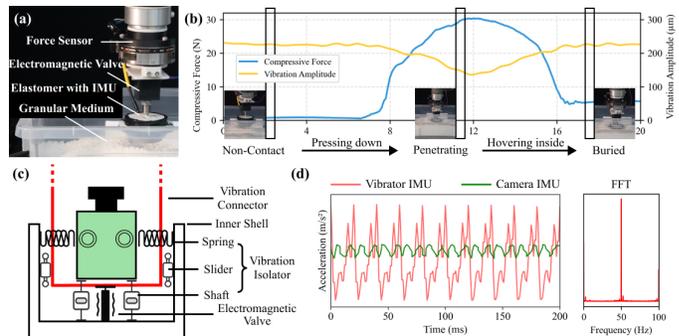

Fig. 13. Hardware validation of the onboard vibration system. (a) Test setup to measure vibration amplitude versus compressive force. (b) Evolution of compressive force and vibration amplitude during penetration into granular media. (c) Structural schematic of the vibration isolation mechanism. (d) IMU measurements for isolation and FFT analysis.

of horizontal vibration is small, it effectively compensates when vertical vibration performance is suboptimal. Moreover, horizontal vibration helps release stress accumulated from vertical oscillation, contributing to structural stability mechanically. In summary, the combined 100 Hz horizontal and 50 Hz vertical vibrations yield the best overall performance.

5) Elastomer Viscoelasticity: The performance peak observed in our experiments is closely related to the material properties of the PDMS elastomer. Theoretically, PDMS is a viscoelastic material whose mechanical behavior are primarily influenced by cross-linking ratio [56] and material thickness [57]. It exhibits time-delayed responses and energy dissipation, and in vibrational environments, the strain-rate effect further reduces the effective elastic range at higher frequencies [58]. Although the transient response of elastomers is studied in acoustics (typically >1000 Hz) [57] and biomedical applications (around 1 Hz) [59], experimental data in the mechanical vibration range (10–100 Hz) remains scarce. Therefore, we adopted a modeling approach to illustrate the general trend. We used the Kelvin-Voigt model to describe this behavior, whose constitutive equation is:

$$\sigma(t) = E\varepsilon(t) + \eta\frac{\mathrm{d}\varepsilon(t)}{\mathrm{d}t}. \tag{13}$$

This model shows the elastomer acts as a first-order low-pass filter. At lower frequencies ($\omega \to 0$), it exhibits a near-ideal elastic response (gain $\to 1/E$), but at higher frequencies ($\omega \to \infty$), viscous damping dominates, causing significant amplitude reduction and phase lag. The characterization result in Fig. 10(d), showing a clear MSNR decline in the 50–120 Hz range, aligns with this model.

To confirm this, we fabricated and tested three elastomers with different hardness levels (Fig. 12(a)). The results in Fig. 12(b) show that as elastomer hardness increases, its elastic response is enhanced, and the MSNR peak frequency also rises. This confirms the relationship between material

viscoelasticity and the damping behavior transition. The elastomer with a hardness of approximately 20 A (17:1 PDMS formulation) consistently delivered superior performance, balancing signal inconsistency at lower frequencies and energy loss at higher frequencies. The softer elastomer failed to capture texture details, while the harder one required higher contact forces, degrading performance.

V. EXPERIMENTS

In this section, we first validate vibration designs in SWTac and the sensor performance on core tactile tasks: contact surface estimation, force estimation, and stone classification. We then evaluate the integrated SandWorm robot, assessing its penetration, autonomous locomotion, pipeline dredging, and subsurface exploration capabilities. Finally, we demonstrate the system's real-world applicability through locomotion and perception field tests.

A. Vibration System Validation

In this experiment, we validated the vibration design of SWTac. First, we tested the performance of the onboard electromagnetic valve to ensure that its active vibration remains effective under actual compressive loads. As shown in Fig. 13(a), we constructed a test platform with a force sensor to measure compressive load and an external IMU attached to the elastomer to measure vibration. We examined the amplitude-frequency characteristics under different contact force conditions. The results indicate that the vibration
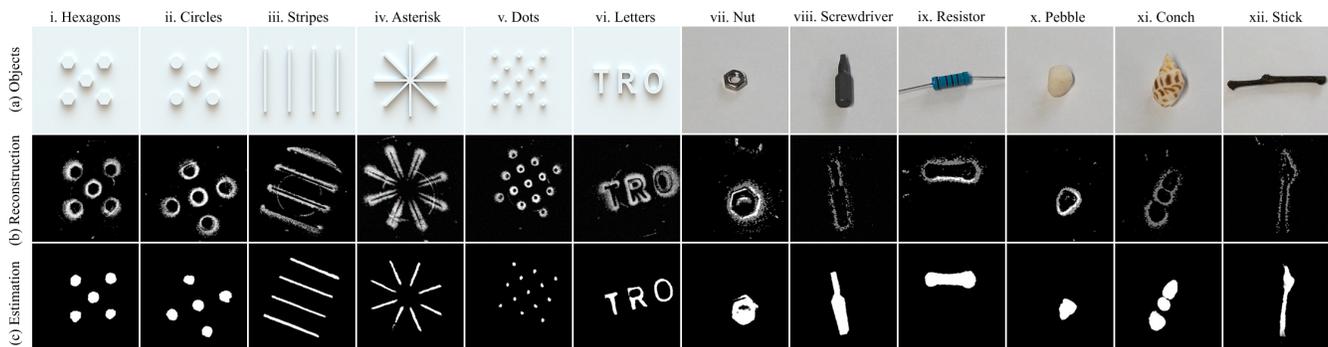
Fig. 14. Contact surface estimation. (a) Texture boards and real-world objects. (b) Raw accumulated grayscale images. (c) Estimation result by U-Net.

frequency consistently remained stable at 50 Hz, regardless of the contact force. Fig. 13(b) shows that as the elastomer is pressed and hovered within the granular medium, the vibration amplitude decreases with increasing contact pressure. However, even under a high compressive load of 30 N, the electromagnetic valve maintained an amplitude of 134 μm. While this reduces the MSNR by approximately 20% from its maximum, the resulting signal remains sufficiently clear for effective perception. We further confirmed that under other normal pressure conditions, such as during rotation and in different granular media, the amplitude also remains well within the ideal operating range.

Furthermore, we verified the effectiveness of the hardware-level vibration-isolation mechanism, which is crucial for stable imaging. As shown in Fig. 13(c), the electromagnetic valve actuates the elastomer through the vibration connector, while the event camera is mounted to the inner shell via several vibration isolators. To validate this structure, we modified the SWTac to attach an IMU to the vibration connector, and thereby compare its data to the event camera's internal IMU. The results in Fig. 13(d) confirm an 83% reduction in acceleration transmitted to the camera, proving the hardware decoupling significantly attenuates vibration. Additionally, the FFT analysis shows the vibration frequency remains precisely locked at 50 Hz with an error of less than 0.2%. This stability originates from the MOSFET switch controlling the electromagnetic valve.

To this end, based on the characterization results and this hardware validation, we established the sensor configuration used for all subsequent experiments: a 17:1 PDMS elastomer, a 50 Hz vertical vibration from the electromagnetic valve, a 100 Hz horizontal vibration, and a mid-level event sensitivity setting.

B. Contact Surface Estimation

Motivated by these intuitive asymmetric edge features, we adopted a U-Net architecture [60] to automatically recover the contact surface from raw accumulated edge images. This architecture integrates multi-scale context while preserving boundary detail through skip connections. It is well suited to delineate sharp contact boundaries and to suppress blur introduced by lateral elastomer deformation. We adopt a 4-level U-Net, which first maps a single-channel grayscale input

to a single-channel mask. Each encoder stage uses two 3×3 convolutions (BN+ReLU), followed by 2×2 max pooling; the bottleneck repeats the double 3×3 convolutions. The decoder upsamples with 2×2 transposed convolutions and skip connections, each followed by two 3×3 convolutions (batch normalization and ReLU). A final 1×1 convolution with a sigmoid produces the mask.

For the dataset, we used 12 objects shown in Fig. 14(a), including six 3D-printed texture boards and six real-world objects. A total of 300 reconstructed tactile images are collected with manual annotations of the contact region. The accumulated images are shown in Fig. 14(b). Moreover, Data augmentation expanded the set to 3000 images via stochastic transforms. For cross-category evaluation, we adopted category-holdout splits: six categories for training and all for testing, with 1000 training and 1000 testing images per split.

For the training process, input images were resized to 256×256. The model is trained with an ℓ_1 loss and Adam optimizer [61] (learning rate 1×10^{-4}), using a batch size of 64 for 200 epochs. The training process ran on an NVIDIA 5090 GPU with an AMD R9-9950X CPU. Three metrics were used to evaluate the estimation result: structural similarity index measure (SSIM), intersection over union (IoU), and RMSE. SSIM measures the perceptual similarity between reconstruction and ground truth, IoU quantifies the overlap between predicted and true contact regions, and RMSE indicates the average pixel-wise error between the estimation and ground truth. We trained on five randomly selected category-holdout splits and compared models with and without the IMU-guided temporal filter; results are reported in Table V.

TABLE V
QUANTITATIVE RESULT OF CONTACT SURFACE ESTIMATION

	SSIM ↑	IoU ↑	RMSE ↓
Filtered	**0.9688 ± 0.0004**	**0.8104 ± 0.0052**	**0.0693 ± 0.0011**
Raw	0.9684 ± 0.0043	0.6965 ± 0.0077	0.0945 ± 0.0016

Across all splits, U-Net delivers strong estimation performance and IMU-guided filtering consistently outperforms raw inputs. The increase in IoU shows cleaner region delineation and fewer boundary errors. The decrease in RMSE reflects smaller pixel-wise deviations. SSIM remains comparable, which suggests the U-Net already captures global struc-
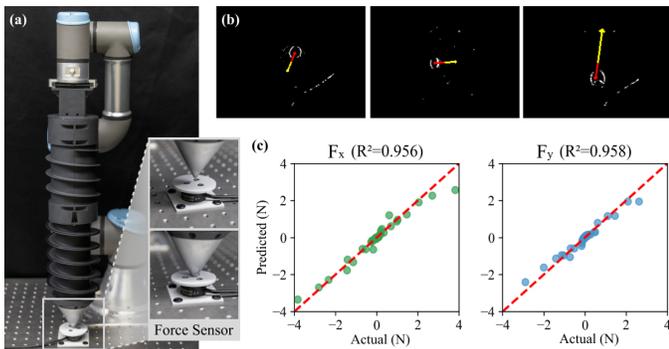
Fig. 15. Shear force estimation. (a) Calibration platform highlighting the contact between the elastomer and the force sensor. (b) Illustration of force estimation results. (c) Regression results of predicted and actual forces.

Fig. 16. Stone classification. (a) Sample images of different types of stones. (b) Confusion matrix of classification results.

ture. These gains indicate that filtering suppresses vibration-induced jitters and stabilizes event density, enabling the U-Net to sharpen boundaries and reject shadowing from lateral elastomer deformation. Fig. 14(c) intuitively shows that the predicted masks align closely with the true contact regions, capturing fine contours and small gaps while minimizing spurious halos. For the limitation, the high textural density near the asterisk pattern center leads to insufficient elastomer deformation, making it difficult to accurately reconstruct the modelled geometry. Together, these results validate the proposed algorithm and demonstrate that SWTac provides reliable visuotactile output for marker-free contact geometry recovery.

C. Force Estimation

In this experiment, we employ a conical elastomer to validate SWTac's capability to provide force information for downstream tasks. The force can be estimated by tracking the displacement at the tip [36]–[38]. We focus on shear measurements because the sensor's active vibration is primarily in the normal direction, making the normal force strongly oscillatory and of limited practical value. Though it can be estimated robustly via short-time averaging, this comes at the cost of a reduced effective sampling rate. We developed the calibration platform shown in Fig. 15(a). By precisely controlling the sensor's lateral offset with a robotic arm, we recorded event streams corresponding to different shear forces. As the offset increases, the tip eventually slides on the force sensor due to insufficient friction, defining the onset of the maximum shear force. A total of 36 data segments were collected for this analysis.

Due to the simplicity of the target geometry, we track the featured tip from the denoised grayscale images using Canny edge detection [62] with tuned spatiotemporal constraints. Using the undeformed elastomer tip as the baseline reference, we compute the minimum enclosing circle of the tracked tip and record its central geometry (x, y) relative to this reference. With the collected dataset, we trained regression models to estimate the force vector (F_x, F_y) from the extracted geometric features. An auxiliary feature r is calculated as the Euclidean distance from the feature center to the reference center. For shear forces F_x and F_y, we adopted input features
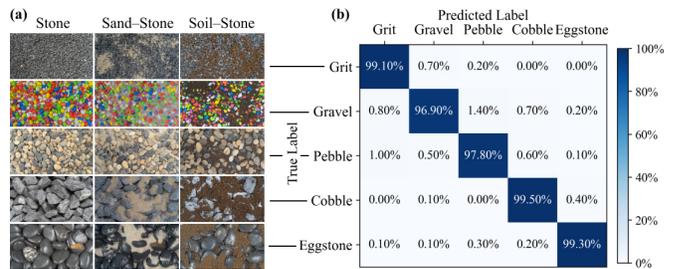
$[x, y, r]$ and used the Random Forest [63] regressor for its ability to capture nonlinear dependencies and its robustness on small, noisy datasets. Evaluation on the dataset showed high accuracy, as illustrated in Fig. 15(c), with R^2 scores exceeding 0.95 for shear forces and mean average error (MAE) below 0.15 N. For each frame, the predicted force vector (F_x, F_y) was overlaid on the image as shown in Fig. 15(b). The red arrow represents the displacement of the featured tip from the reference position, indicating direction, while the yellow arrow shows the shear force direction and magnitude. For display purposes, the force direction was inverted.

D. Stone Classification

Accurate classification of stone types, which exhibit unique granularity and surface textures, is essential for subsurface exploration. This experiment tests SWTac's ability of texture and validates its robustness in realistic downstream applications. We collected the tactile data for five stone types in the real world: grit, gravel, pebble, cobble, and eggstone. Data was gathered under three conditions: stone only, sand–stone composites, and soil–stone composites, as shown in Fig. 16(a). The IMU-guided filter was applied to the signals after removing non-contact intervals with reconstruction rate at 500 Hz. Data augmentation was applied to the dataset, yielding a total of approximately 10,000 images.

For classification, we adopted a ResNet-18 model pre-trained on ImageNet [64]. We employed a transfer learning strategy, freezing the stem modules to preserve low-level features. These modules contain the initial convolution, batch normalization, and max pooling. The remaining residual stages were fine-tuned. The final fully connected layer was replaced with a dropout layer and a new linear layer with five outputs. The dataset was split into training and test sets at a 1:1 ratio. The model was trained for 50 epochs using an Adam optimizer with a learning rate of 10^{-4}, an ℓ_2 decay of 10^{-4}, and a batch size of 32 with cross-entropy loss. Training was conducted on an NVIDIA 5090 GPU with an AMD R9-9950X CPU.

The resulting confusion matrix is shown in Fig. 16(b). The model achieved an overall accuracy of 98%, successfully classifying the stone types even in these unstructured, composite environments. This high robustness is attributed to the sensor's active vibration mechanism. The high-frequency vibrations help dislodge fine particles like sand and soil from the stone's surface, reducing interference and allowing the sensor to capture the underlying tactile signature.
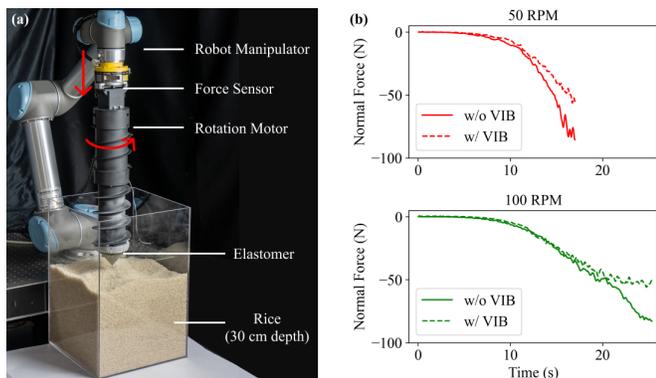
Fig. 17. Benchmark of vibration-assisted penetration. (a) Experimental setup. (b) Force sensor readings under static and vibrational excitation at different rotational speeds during penetration.

E. Penetration Benchmark

In this experiment, we quantify the contributions of the screw thread and active vibration to the penetration performance of the SandWorm robot. To create a standardized test, we constructed a test platform, as shown in Fig. 17(a), using rice as the benchmark medium. The rice was piled approximately 300 mm high in a container. SWTac was fixed to a robotic arm via a high-range force sensor, with the brushless motor providing rotational motion. A soft dust-proof mesh was installed around the elastomer to prevent particle ingress, a measure maintained in all subsequent experiments. The robotic arm descended at a constant speed of 5 mm/s. By varying the rotational speed, we compared the effect of vibration on penetration performance.

The results in Fig. 17(b) show that higher rotational speeds facilitated easier penetration. Crucially, the presence of vibration consistently led to lower penetration resistance at the same depth across all rotational speeds. At 100 RPM, the sensor penetrated beyond 200 mm with relatively stable force readings. Furthermore, the beneficial effect of vibration was most pronounced at lower rotational speeds. These findings confirm that active vibration significantly enhances penetration efficiency in granular media, validating the effectiveness of the sensor's structural design.

F. Pipeline Inspection

In this experiment, we first characterized SandWorm's linear locomotion in a 1 m long, 200 mm inner diameter acrylic pipe (Fig. 18(a)). We tested four motion configurations over a 600 mm distance: rotational speeds of 60 and 90 RPM, each with and without the 30 mm pushrod reciprocation. The results are summarized in Table VI. It shows that the peristaltic motion by pushrod reciprocation reduces traversal time by up to 62%. This improvement stems from the alternation of extension and retraction by the pushrod, which overcomes static friction more effectively than screw actuation alone. This synergy yielded a maximum straight-line speed of 12.5 mm/s, confirming that integrated pushrod motion is critical for efficient locomotion.

Next, we evaluated the steering performance of SandWorm in a 15° pipe bend. Fig. 18(b) shows the starting and finishing

Fig. 18. Performance evaluation of the autonomous SandWorm robot in confined environments. (a) Linear locomotion within a straight pipe. (b) Steering inside the pipe. (c) Steering at wall intersections. (d) Traversal through a curved pipe segment.

TABLE VI
TRAVERSAL TIME OF SANDWORM IN THE PIPELINE

Configuration	60 RPM	90 RPM
w/o Pushrod Reciprocation	140 s	127 s
w/ Pushrod Reciprocation	62 s	48 s

configurations. For propulsion, we employed a rotational speed of 90 RPM combined with pushrod reciprocation. The steering is enabled by force estimation, which triggers the servo. The conical elastomer was mounted on the tip and calibrated for shear force sensing. When the elastomer tip contacts the inner wall of the pipe bends, the induced deflection generates events captured by the event camera (Fig. 18(b)ii). We set a threshold on the shear force change to trigger the servo at 1 Hz, producing a 30° rotation of the servo motor per trigger; successive triggers accumulate to achieve larger steering angles. Since the robot and sensor rotate, the detected force direction must be transformed into the world frame. To achieve this with low computational overhead, we use a simple quantization method [65] to correct the force vector based on the motor's angle, mapping the result to a left or right steering command. Using this approach, the robot successfully navigated the bend.

In addition, we evaluated steering at wall intersections, as shown in Fig. 18(c). SandWorm travels along a wall inclined at 70° to the horizontal plane and then contacts a vertical wall that intersects its direction of motion at a 70° angle. The algorithm and steering mode are identical to the last one. Experimentally, the steering manoeuvre unfolds through four observed phases: i. nominal forward motion along the initial
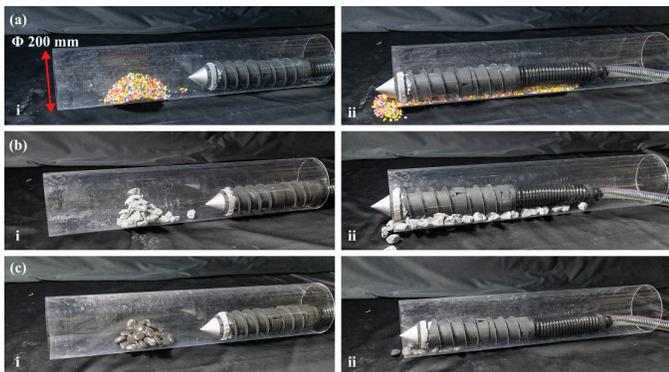
Fig. 19. Pipeline dredging experiments with the autonomous SandWorm robot. (a) Gravel blockage. (b) Cobble blockage. (c) Eggstone blockage. i. image before pipe dredging, ii. image after pipe dredging.

wall; ii. elastomer tip contacting with the new wall, triggering the servo; iii. reorientation toward the new wall followed by renewed contact detection; and iv. cyclic actuation and contact checks until the elastomer tip no longer contacts the wall. Once the servo returns to its neutral position, the robot resumes forward motion. Aided by its rotational motion, SandWorm would naturally align with the new wall, thereby completing the large-angle steering.

Finally, we evaluated extreme locomotion in a more challenging 150 mm diameter PVC elbow with a 2000 mm curvature radius, as shown in Fig. 18(d). The material and tightened diameter introduced significantly higher friction. These conditions caused the pushrod to stall. To mitigate this, we reduced the motor speed to 30 RPM and relied on continuous low-speed rotation, which prevented shock loads. The turn was gradual, so the steering servo remained neutral. Despite the severe friction, the robot successfully traversed the bend, maintaining forward motion at 12% of its maximum speed. This result demonstrates the mechanical robustness and compliance of the design under extreme frictional loads.

G. Pipeline Dredging

Clearing obstructions serves as a capstone experimental validation, demonstrating the integrated capabilities of the SandWorm robot. It combines the previously characterized locomotion, perception, and mechanical design. We tested the robot's dredging performance against three blockage types: gravel, cobble, and eggstone as shown in Fig. 19. Each blockage spanned 200 mm and occupied approximately half the pipe's cross-section. SandWorm advanced while actively clearing the obstruction, successfully traversing the 600 mm path in 84 s, 90 s, and 81 s for each case, respectively. Compared to the 48 s unobstructed run, the speed decreased by approximately 44%, yet the robot consistently cleared and traversed all three blockage types.

SandWorm's high-torque drive and threaded housing fragmented the blockage and then redistributed debris along the pipe, like a tunnel-boring machine. The helical shell ensured uniform debris spreading and even backward transport. During dredging, SWTac captures the event stream continuously,
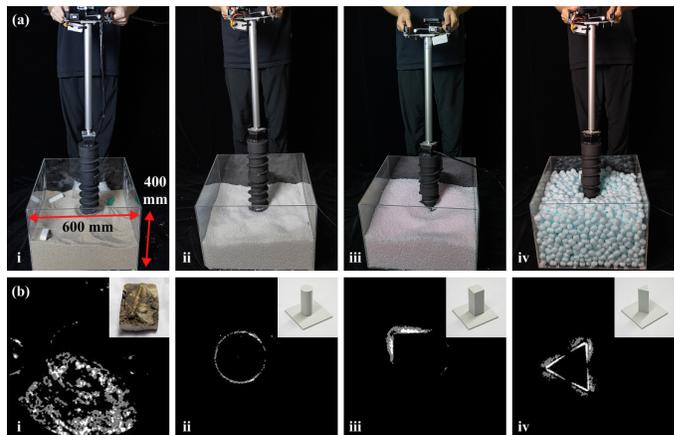

Fig. 20. Subsurface exploration in granular media by SandWorm. (a) Manual experiments in: i. sand, ii. TPE, iii. EPP, iv. EPE. (b) Shapes and imaging of detected objects: i. trilobite fossil, ii. cylinder, iii. cuboid, iv. triangular prism.
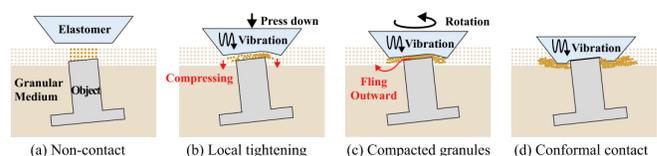

Fig. 21. Sequence of elastomer-object contact in granular media, showing (a) no contact, (b) blurring from initial pressing, (c) granule compaction, and (d) granule evacuation via vibration and rotation to reveal clear object edges.

which is processed by our stone classification network. A majority vote over sequential frames enabled robust and accurate identification of blockage material throughout the traversal. This confirms that SandWorm can not only physically dredge a pipeline but also simultaneously perceive its environment, validating the complete system integration in a challenging, application-focused scenario.

H. Subsurface Exploration

For subsurface exploration, we use the manual version of SandWorm, replacing the tail module with a holding pushrod for downward force while retaining rotational drilling, as illustrated in Fig. 20(a). The frustoconical elastomer was chosen for texture perception at the tip. The event stream was reconstructed, passed through a $3{\times}3$ median filter, and visualized in real-time at 200 Hz. A smartphone interface provided communication and allows the operator to control the rotation speed and pushrod actuation.

We evaluated the object-searching performance in a $600{\times}600{\times}400$ mm^3 test box with four media: sand, thermoplastic elastomer (TPE), expanded polypropylene (EPP), and expanded polyethylene (EPE), with respective densities of approximately 2000, 700, 100, and 10 kg/m^3. The key to this exploration is distinguishing the target object from the surrounding granular media by volume, density, and texture. As the tip penetrates, the active vibration mechanism gradually clears away granules from the object's surface. This allows the tip to make direct contact, revealing the object's true shape and texture rather than the granular pattern. The process is illustrated in Fig. 21.
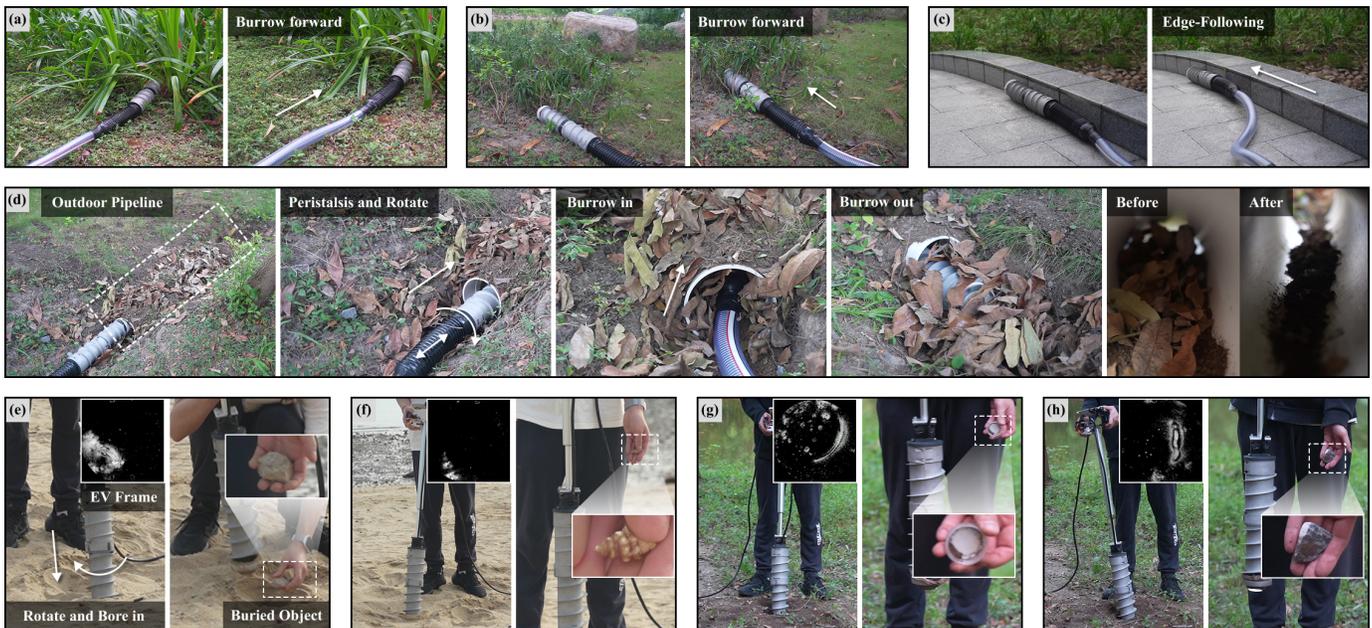
Fig. 22. Field experiments demonstrating SandWorm's capabilities in real-world environments. (a) Moving forward in grass. (b) Moving forward in a bush. (c) Moving forward on a cement road. (d) Dredging in a blocked outdoor pipeline. (e)(f) Subsurface exploration experiment on the beach. (g)(h) Subsurface exploration experiment in composite garden soil. Full experimental footage is available in the supplementary video.

In 40 trials, operators achieved a 90% success rate (36/40) in locating buried objects within 120 s. Sample imaging results are shown in Fig. 20(b). As predicted by the mechanism, searching in dense sand and TPE was slower; all four trials exceeding 120 s occurred in these media. These failures in dense media occurred because high inertial torque caused the robot body to pivot, shifting the contact point to the elastomer's conical side. This resulted in reduced localization accuracy, which can be mitigated by reducing rotation speed, though at the cost of slower penetration.

I. Field Validation

To validate SandWorm's real-world performance, we conducted a final series of experiments in unstructured outdoor environments, as shown in Fig. 22. These tests were designed to evaluate the system's integrated performance in complex, unpredictable conditions, where engineering improvements such as SandWorm's metal shell and the elastomer's dust-proof mesh proved crucial for durability.

First, we assessed SandWorm's locomotive performance on challenging natural terrains. SandWorm successfully traversed dense grass (Fig. 22(a)), tangled bushes (Fig. 22(b)), and hard cement surfaces (Fig. 22(c)), demonstrating the robustness of its rotational-peristaltic locomotion. We then validated its applications in pipeline dredging. The robot was deployed in a blocked outdoor pipe, shown in Fig. 22(d). Upon entering, SandWorm continuously burrowed in, actively clearing internal blockages composed of mud, leaves, and gravel. Throughout this process, SWTac indicated no prohibitive shear forces, allowing the robot to autonomously continue the clearing operation and traverse the pipe.

For subsurface exploration, we conducted numerous trials in complex, non-uniform media. The operator uses SandWorm

to bore in at different locations until the target is detected in the reconstructed tactile images, as shown in Fig. 22(e). We show four representative successes in Fig. 22(e-h). SandWorm successfully located diverse buried objects, including a stone and a conch on a beach, as well as a bottle cap and another stone in composite garden soil. These field tests confirm that SandWorm's integrated locomotion and active perception are effective and generalizable for its target applications.

VI. Discussion

The SandWorm robot addresses key challenges in granular media environments that traditional perception and locomotion methods face, such as low resolution, poor robustness, and vulnerability to strong vibration. By integrating a bio-inspired rotational-peristaltic locomotion mechanism with a novel event-based visuotactile sensor, SandWorm achieves efficient movement and high-resolution perception, demonstrating clear advantages in stability and adaptability over conventional approaches. Ultimately, this work demonstrates that vibration can be engineered into a perception advantage when hardware and algorithms are co-designed from first principles.

Scientific Contributions. This work makes three core contributions: (1) We demonstrate that active vibration enables event cameras to effectively capture static contact. By vibrating the elastomer tip, SWTac converts static contact into detectable dynamic events, enabling pixel-level tactile imaging at 1000 Hz without motion blur. This approach fundamentally expands event-based vision beyond dynamic scenes. (2) We introduce hardware-level perception-actuation integration, where the same vibration drives both imaging and material penetration, while a dedicated isolation mechanism decouples

and protects the sensor. (3) We propose a bio-inspired locomotive structure combining rotation and peristalsis, which mimics marine polychaetes to achieve up to 62% faster traversal than screw actuation alone.

Algorithmic Innovations. This work provides innovations in two key areas: (1) While static tactile imaging by event camera is enabled by our hardware design, we further optimized the temporal continuity of event-based perception through algorithmic solutions. We proposed IMU-guided temporal filter that exploits the correlation between elastomer acceleration and imaging quality, improving MSNR by 24% in average with 1 ms latency. (2) We developed algorithms for key downstream applications, including recovering full contact surfaces by leveraging asymmetric edge deformation with FEA, force estimation, and material classification.

Engineering Insights. This work illuminates several lessons for extreme-environment robotics: (1) Event cameras excel when vibration is harnessed as signal, not noise. SWTac transforms vibration from an interference source into an advantage, extending the application of event cameras to static tactile perception. (2) Multimodal fusion of event-based vision with IMU are essential for maintaining signal fidelity from high-sample-rate sensors under high mechanical loads, applicable to any contact-rich task.

Limitations and Future Work. Despite these advances, particle adhesion to the elastomer surface may impair texture detection in sticky soils. Future work will explore tuning the elastomer's surface properties via hydrophobic coatings or micro-texturing, as well as active cleaning mechanisms (e.g., air jets, ultrasonic agitation). Additionally, extending contact surface estimation to 3D reconstruction would enable richer geometric reasoning, which can be adopted for real-time adaptive control to improve autonomous performance significantly. These steps represent a clear path toward reliable robotic perception for advanced applications in disaster rescue, archaeological excavation, and planetary exploration.

VII. CONCLUSION

In this paper, we introduce SandWorm, a peristalsis-augmented screw-actuated snakelike robot, with SWTac, an event-based visuotactile sensor integrated with active elastomer vibration. By mechanically isolating the event camera, SWTac enables consistent tactile imaging in pixel-level at 1000 Hz, overcoming event cameras' intrinsic limitation of capturing only dynamic changes. For the algorithm, we proposed the IMU-guided temporal filter to enhance the consistency of the asynchronous event stream by approximately 24%. Moreover, we verified the asymmetric edge feature by FEA and adopted the U-Net model for contact surface estimation, as well as systematically characterized sensor performance. Experimentally, SWTac provided precise force estimation (0.15 N), texture recognition (0.2 mm resolution), and stone classification (98% accuracy). The SandWorm robot demonstrated effective autonomous locomotion, achieving straight runs (12.5 mm/s), turns, high-curvature navigation. Field trials in real-world scenarios, including pipeline dredging, subsurface exploration in beach and composite soil, further demonstrate the system's practical effectiveness.
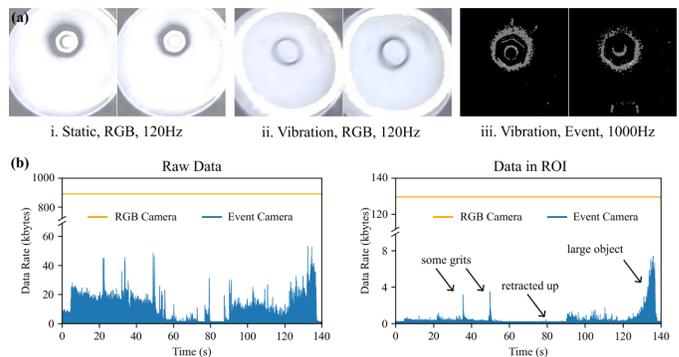

Fig. 23. Event Camera vs. RGB Camera. (a) Intuitive illustration of frames. (b) Data rate comparison during the field test on the beach.

APPENDIX A
DERIVATION OF LOCOMOTION

This appendix provides the mathematical derivations for the locomotion components referenced in Section III-C. First, the screw mechanism converts the external shell's rotation θ into an axial displacement l along the robot's intrinsic axis. This relationship is given by:

$$l = \frac{p}{2\pi}\theta, \tag{14}$$

where p is the pitch of the screw.

Second, we derive the effective gravitationally resolved driving component, F_{propel}. We consider the robot on an inclined plane with inclination α, mass m, and under gravitational acceleration g. An axial displacement $\mathrm{d}l$ induces a vertical height change $\mathrm{d}h = \mathrm{d}l \sin\alpha$. The corresponding change in gravitational potential energy $\mathrm{d}U$ is:

$$\mathrm{d}U = mg\mathrm{d}h = mg\sin\alpha\mathrm{d}l. \tag{15}$$

F_{propel} is defined as the potential energy gradient along the robot's intrinsic axis. Taking the derivative of U with respect to l from Eqn. (15) yields this gravitational component, which is used in the net axial force analysis in Eqn. (1) and Eqn. (2):

$$F_{\text{propel}} = \frac{\mathrm{d}U}{\mathrm{d}l} = mg\sin\alpha. \tag{16}$$

APPENDIX B
CAMERA COMPARISON

We conducted a direct comparison with an RGB camera (OV2710) at 120 Hz frame rate with exposure time fixed at 1.6 ms, as shown in Fig. 23. While the RGB camera produces clear images in static contact, it suffers from significant motion blur and distortion underthe active vibration required for our perception method. Although further reducing the exposure could mitigate the blurring, it would lead to underexposure and electronic noise, degrading the signal quality. In contrast, the event camera robustly remains sharp even when reconstructed at 1000 Hz under high-frequency vibrations. Its asynchronous pixel-level change detection enables a response latency as low as 15 μs [49], serving as an effectively ultra-short exposure. This surpasses traditional photometric integration in sensitivity
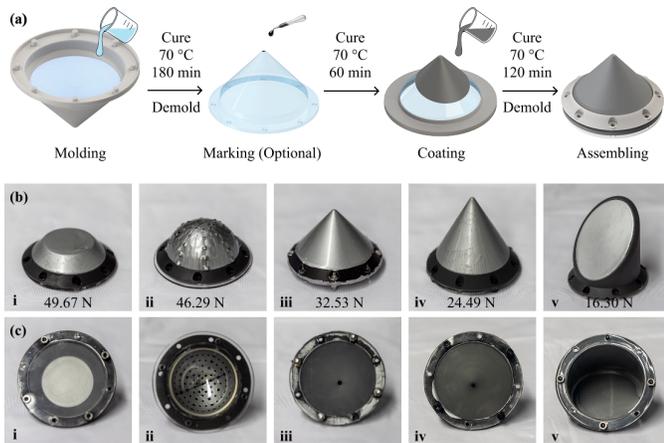
Fig. 24. Fabrication of different elastomers. (a) Schematic of the elastomer fabrication procedure. (b) Photograph of the moulded elastomer, annotated with maximum contact force values. (c) Interior (imaging) side of the elastomer.

to rapid illumination changes, ensuring robust tactile perception with high temporal resolution.

We also performed a quantitative analysis of the data rate [37]. Using an event stream captured during a field test on the beach for comparison, we found that even at the moment of peak event generation while identifying an object, the event camera is vastly more data-efficient. This data efficiency is critical for reducing the computational and transmission loads on a computationally constrained mobile robot.

APPENDIX C
ELASTOMER DESIGN

The elastomer fabrication procedure is illustrated in Fig. 24(a). PDMS base and curing agent are mixed at the target ratio, degassed under vacuum, cast into a customized mold, cured at 70 °C for 180 min, and then demolded. Next, markers are deposited by applying a thin PDMS–carbon powder composite, which is cured at 70 °C for 60 min. A reflective layer of PDMS doped with silver powder is then spin-coated over the elastomer, using a temporary insert to fill any gaps, and subsequently cured at 70 °C for 120 min. Finally, the elastomer is assembled with an acrylic substrate and mechanical latch (P1 and P3 in Fig. 2(e)) to protect the transparent bottom surface from contamination. To achieve different hardness levels, PDMS samples were cast with 10:1 and 17:1 base-to-curing-agent mass ratios, yielding approximate Shore hardness values of 40 A and 20 A, respectively. A 10 A hardness is realized using silicone (Ecoflex™ 00-10, Smooth-On Inc.). To evaluate long-term durability, we conducted a 15,000-cycle rotational wear test. Based on the results, we enhanced the durability by adding a thin, abrasion-resistant TPU film to the elastomer surface.

Different shapes of elastomer can be used for different functions [66]. To enhance sensor performance under different contact conditions, five elastomer geometries were fabricated. The external surfaces and the internal marker distributions are shown in Fig. 24(b) and (c), respectively. These geometries include hemispherical designs, which incorporate a dense array

of markers to achieve a wide dynamic range of force perception, and obliquely cut cylindrical geometries, which combine extended lateral surfaces with a single inclined marker line designed to support efficient drilling in granular media. Conical tips feature a single central marker and tunable aspect ratios to provide graded force sensitivity. Finally, frustoconical designs provide a flat terminal surface and broad contact area, enhancing texture discrimination. Furthermore, we experimentally validated these elastomer geometries by testing the maximum normal force experienced during penetration into and retraction from the same granular medium under identical conditions. The resulting values are annotated below each geometry in Fig. 24(b).

REFERENCES

[1] B. Kou, Y. Cao, J. Li, C. Xia, Z. Li, H. Dong, A. Zhang, J. Zhang, W. Kob, and Y. Wang, "Granular materials flow like complex fluids," *Nature*, vol. 551, no. 7680, pp. 360–363, 2017.

[2] S. J. Antony, W. Hoyle, and Y. Ding, *Granular Materials: Fundamental and Applications*. The Royal Society of Chemistry, 03 2004.

[3] A. Katterfeld and C. Wensrich, "Understanding granular media: from fundamentals and simulations to industrial application," *Granular Matter*, vol. 19, no. 4, p. 83, 2017.

[4] D. Lattanzi and G. Miller, "Review of robotic infrastructure inspection systems," *Journal of Infrastructure Systems*, vol. 23, no. 3, p. 04017004, 2017.

[5] S. K. Pal, "Granular mining and big data analytics: Rough models and challenges," *Proceedings of the National Academy of Sciences, India Section A: Physical Sciences*, vol. 90, no. 2, pp. 193–208, 2020.

[6] A. H. Reddy, B. Kalyan, and C. S. Murthy, "Mine rescue robot system – a review," *Procedia Earth and Planetary Science*, vol. 11, pp. 457–462, 2015.

[7] T. S. Vaquero, G. Daddi, R. Thakker, and et al., "EELS: Autonomous snake-like robot with task and motion planning capabilities for ice world exploration," *Science Robotics*, vol. 9, no. 88, p. eadh8332, 2024.

[8] R. Patel, R. Ouyang, B. Romero, and E. Adelson, "Digger Finger: GelSight tactile sensor for object identification inside granular media," in *Proc. International Symposium on Experimental Robotics (ISER)*, Floriana, Malta, 2021.

[9] X. Xiao and R. Murphy, "A review on snake robot testbeds in granular and restricted maneuverability spaces," *Robotics and Autonomous Systems*, vol. 110, pp. 160–172, 2018.

[10] F. Pacheco-Vázquez and J. Ruiz-Suárez, "Cooperative dynamics in the penetration of a group of intruders in a granular medium," *Nature Communications*, vol. 1, no. 1, p. 123, 2010.

[11] W. Kang, Y. Feng, C. Liu, and R. Blumenfeld, "Archimedes' law explains penetration of solids into granular media," *Nature Communications*, vol. 9, no. 1, p. 1101, 2018.

[12] S. Li, H. Yu, G. Pan, H. Tang, J. Zhang, L. Ye, X.-P. Zhang, and W. Ding, "M³tac: A multispectral multimodal visuotactile sensor with beyond-human sensory capabilities," *IEEE Transactions on Robotics*, vol. 40, pp. 4484–4503, 2024.

[13] R. S. Dahiya, G. Metta, M. Valle, and G. Sandini, "Tactile sensing—from humans to humanoids," *IEEE Transactions on Robotics*, vol. 26, no. 1, pp. 1–20, 2010.

[14] A. Ghasemloonia, D. Geoff Rideout, and S. D. Butt, "A review of drillstring vibration modeling and suppression methods," *Journal of Petroleum Science and Engineering*, vol. 131, pp. 150–164, 2015.

[15] B. Saldivar, I. Boussaada, H. Mounier, S. Mondié, and S. Niculescu, "An overview on the modeling of oilwell drilling vibrations," *IFAC Proceedings Volumes*, vol. 47, no. 3, pp. 5169–5174, 2014.

[16] S. Li, Z. Wang, C. Wu, and et al., "When vision meets touch: A contemporary review for visuotactile sensors from the signal processing perspective," *IEEE Journal of Selected Topics in Signal Processing*, vol. 18, no. 3, pp. 267–287, 2024.

[17] R. Dash and B. Majhi, "Motion blur parameters estimation for image restoration," *Optik*, vol. 125, no. 5, pp. 1634–1640, 2014.

[18] L. Pan, C. Scheerlinck, X. Yu, R. Hartley, M. Liu, and Y. Dai, "Bringing a blurry frame alive at high frame-rate with an event camera," in *Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Long Beach, CA, USA, 2019.

[19] G. Gallego, T. Delbruck, G. Orchard, C. Bartolozzi, B. Taba, A. Censi, S. Leutenegger, A. Davison, J. Conradt, K. Daniilidis, and D. Scaramuzza, "Event-based vision: A survey," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 44, pp. 154–180, 2022.

[20] B. He, Z. Wang, Y. Zhou, J. Chen, C. D. Singh, H. Li, Y. Gao, S. Shen, K. Wang, Y. Cao, C. Xu, Y. Aloimonos, F. Gao, and C. Fermüller, "Microsaccade-inspired event camera for robotics," *Science Robotics*, vol. 9, no. 90, p. eadj8124, 2024.

[21] A. B. Tzetlin and A. V. Filippova, "Muscular system in polychaetes (annelida)," *Hydrobiologia*, vol. 535, no. 1, pp. 113–126, 2005.

[22] S. Grill and K. Dorgan, "Burrowing by small polychaetes - mechanics, behavior and muscle structure of capitella sp," *The Journal of experimental biology*, vol. 218, 2015.

[23] F. Richter, P. V. Gavrilov, H. M. Lam, A. Degani, and M. C. Yip, "ARCSnake: Reconfigurable snakelike robot with archimedean screw propulsion for multidomain mobility," *IEEE Transactions on Robotics*, vol. 38, no. 2, pp. 797–812, 2022.

[24] F. Xue, C. Yao, Y. Yuan, Y. Ge, W. Shi, Z. Zhu, L. Ding, and Z. Jia, "Wheel-terrain contact geometry estimation and interaction analysis using a side-wheel camera over deformable terrains," *IEEE Robotics and Automation Letters*, vol. 8, no. 11, pp. 7639–7646, 2023.

[25] Y.-S. Kwon and B.-J. Yi, "Design and motion planning of a two-module collaborative indoor pipeline inspection robot," *IEEE Transactions on Robotics*, vol. 28, no. 3, pp. 681–696, 2012.

[26] A. Kakogawa, T. Nishimura, and S. Ma, "Designing arm length of a screw drive in-pipe robot for climbing vertically positioned bent pipes," *Robotica*, vol. 34, pp. 306–327, 2016.

[27] P. Nourizadeh, F. J. McFadden, and W. N. Browne, "In situ skid estimation for mobile robots in outdoor environments," *Journal of Field Robotics*, vol. 41, no. 1, pp. 179–194, 2024.

[28] H. Kolvenbach, C. Bärtschi, L. Wellhausen, R. Grandia, and M. Hutter, "Haptic inspection of planetary soils with legged robots," *IEEE Robotics and Automation Letters*, vol. 4, no. 2, pp. 1626–1633, 2019.

[29] Ruppert and Badri-Sprowitz, "FootTile: A rugged foot sensor for force and center of pressure sensing in soft terrain," in *Proc. IEEE International Conference on Robotics and Automation (ICRA)*, Paris, France, 2020.

[30] R. A. Russell, "A ground-penetrating robot for underground chemical source location," in *Proc. IEEE International Conference on Field and Service Robotics (FSR)*, Prot Douglas, Australia, 2005.

[31] R. Cao, D. Galor, A. Kohli, J. L. Yates, and L. Waller, "Noise2Image: Noise-enabled static scene recovery for event cameras," *Optica*, vol. 12, no. 1, pp. 46–55, Jan 2025.

[32] A. Yousefzadeh, G. Orchard, T. Serrano-Gotarredona, and B. Linares-Barranco, "Active perception with dynamic vision sensors. minimum saccades with optimum recognition," *IEEE Transactions on Biomedical Circuits and Systems*, vol. 12, no. 4, pp. 927–939, 2018.

[33] T. Taunyazov, W. Sng, B. Lim, H. H. See, J. Kuan, A. F. Ansari, B. Tee, and H. Soh, "Event-driven visual-tactile sensing and learning for robots," in *Proc. Robotics: Science and Systems (RSS)*, Corvalis, Oregon, USA, 2020.

[34] F. Baghaei Naeini, A. M. AlAli, R. Al-Husari, A. Rigi, M. K. Al-Sharman, D. Makris, and Y. Zweiri, "A novel dynamic-vision-based approach for tactile sensing applications," *IEEE Transactions on Instrumentation and Measurement*, vol. 69, no. 5, pp. 1881–1893, 2020.

[35] D. Yin, S. Lu, J. Yang, Y. Zhang, Z. Dai, D. Nan, B. Cai, S. He, and X. Chen, "GelEvent—a novel high-speed tactile sensor with event camera," *IEEE Transactions on Instrumentation and Measurement*, vol. 74, pp. 1–13, 2025.

[36] D. Mukashev, S. Seitzhan, J. Chumakov, S. Khajikhanov, M. Yergibay, N. Zhaniyar, R. Chibar, A. Mazhitov, M. Rubagotti, and Z. Kappassov, "E-BTS: Event-based tactile sensor for haptic teleoperation in augmented reality," *IEEE Transactions on Robotics*, vol. 41, pp. 450–463, 2025.

[37] N. Funk, E. Helmut, G. Chalvatzaki, R. Calandra, and J. Peters, "Evetac: An event-based optical tactile sensor for robotic manipulation," *IEEE Transactions on Robotics*, vol. 40, pp. 3812–3832, 2024.

[38] 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.

[39] Z. Zahiri, D. F. Laefer, and A. Gowen, "Characterizing building materials using multispectral imagery and LiDAR intensity data," *Journal of Building Engineering*, vol. 44, p. 102603, 2021.

[40] A. Gheibi and A. Hedayat, "Ultrasonic investigation of granular materials subjected to compression and crushing," *Ultrasonics*, vol. 87, pp. 112–125, 2018.

[41] S. Lee, D. K. Woo, and H. Choi, "Contactless estimation of soil moisture using leaky Rayleigh waves and a fully convolutional network," *Vadose Zone Journal*, vol. 22, no. 6, p. e20285, 2023.

[42] R. Stannarius, "Magnetic resonance imaging of granular materials," *Review of Scientific Instruments*, vol. 88, no. 5, p. 051806, 05 2017.

[43] F. Guillard, B. Marks, and I. Einav, "Dynamic X-ray radiography reveals particle size and shape orientation fields during granular flow," *Scientific Reports*, vol. 7, no. 1, p. 8155, 2017.

[44] H. Rebecq, R. Ranftl, V. Koltun, and D. Scaramuzza, "High speed and high dynamic range video with an event camera," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 43, no. 6, pp. 1964–1980, 2021.

[45] C. Branyan and Y. Menguc, "Soft snake robots: Investigating the effects of gait parameters on locomotion in complex terrains," in *Proc. IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, Madrid, Spain, 2018.

[46] S. Scheraga, A. Mohammadi, T. Kim, and S. Baek, "Design of an underactuated peristaltic robot on soft terrain," in *Proc. IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, Las Vegas, NV, USA, 2020.

[47] R. Chen, X. Zhu, Z. Yuan, H. Pu, J. Luo, and Y. Sun, "A bioinspired single actuator-driven soft robot capable of multistrategy locomotion," *IEEE Transactions on Robotics*, vol. 40, pp. 2149–2165, 2024.

[48] T. Dachlika and D. Zarrouk, "Mechanics of locomotion of a double screw crawling robot," *Mechanism and Machine Theory*, vol. 153, p. 104010, 2020.

[49] P. Lichtsteiner, C. Posch, and T. Delbruck, "A 128×128 120 dB 15 μs latency asynchronous temporal contrast vision sensor," *IEEE Journal of Solid-State Circuits*, vol. 43, no. 2, pp. 566–576, 2008.

[50] iniVation AG, "DV-processing," 2025. [Online]. Available: https://dv-processing.inivation.com/

[51] A. Böhm, T. Schneider, B. Belousov, A. Kshirsagar, L. Lin, K. Doerschner, K. Drewing, C. A. Rothkopf, and J. Peters, "What matters for active texture recognition with vision-based tactile sensors," in *2024 IEEE International Conference on Robotics and Automation (ICRA)*, 2024, pp. 15 099–15 105.

[52] Q. Li, O. Kroemer, Z. Su, F. F. Veiga, M. Kaboli, and H. J. Ritter, "A review of tactile information: Perception and action through touch," *IEEE Transactions on Robotics*, vol. 36, no. 6, pp. 1619–1634, 2020.

[53] E. Olsson and D. Jelagin, "A contact model for the normal force between viscoelastic particles in discrete element simulations," *Powder Technology*, vol. 342, pp. 985–991, 2019.

[54] S. Moisio, B. León, P. Korkealaakso, and A. Morales, "Model of tactile sensors using soft contacts and its application in robot grasping simulation," *Robotics and Autonomous Systems*, vol. 61, no. 1, pp. 1–12, 2013.

[55] Y. Ito, Y. Kim, and G. Obinata, "Contact region estimation based on a vision-based tactile sensor using a deformable touchpad," *Sensors*, vol. 14, no. 4, pp. 5805–5822, 2014.

[56] I. Teixeira, I. Castro, V. Carvalho, and et al., "Polydimethylsiloxane mechanical properties: A systematic review," *AIMS Materials Science*, vol. 8, pp. 952–973, 2021.

[57] M. A. Trindade, A. Benjeddou, and R. Ohayon, "Modeling of Frequency-Dependent Viscoelastic Materials for Active-Passive Vibration Damping," *Journal of Vibration and Acoustics*, vol. 122, pp. 169–174, Apr. 2000.

[58] D. Kumar and S. Singh, "Static and dynamic mechanical characterization of polydimethylsiloxane (PDMS) under uniaxial tensile loading," in *Proc. IOP Conference Series: Materials Science and Engineering*, 2022.

[59] J. H. Kim, P. Chhai, and K. Rhee, "Development and characterization of viscoelastic polydimethylsiloxane phantoms for simulating arterial wall motion," *Medical Engineering & Physics*, vol. 91, pp. 12–18, May 2021.

[60] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, Munich, Germany, 2015.

[61] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *Proc. International Conference on Learning Representations (ICLR)*, San Diego, CA, USA, 2015.

[62] J. Canny, "A computational approach to edge detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. PAMI-8, no. 6, pp. 679–698, 1986.

[63] L. Breiman, "Random forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.

[64] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, USA, 2016.

[65] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision*, vol. 60, no. 2, pp. 91–110, 2004.

[66] U. H. Shah, R. Muthusamy, D. Gan, Y. H. Zweiri, and L. Seneviratne, "On the design and development of vision-based tactile sensors," *Journal of Intelligent & Robotic Systems*, vol. 102, pp. 1–27, 2021.

Shoujie Li received the B.Eng. degree in electronic information engineering from the College of Oceanography and Space Informatics, China University of Petroleum, Tsingtao, China, in 2020. He received Ph.D. degree from Tsinghua Shenzhen International Graduate School, Tsinghua University, Shenzhen, China, in 2025. He is working as a research fellow at Nanyang Technological University.

His research interests include tactile perception, grasping, and machine learning. He received the Outstanding Mechanisms and Design Paper Finalists in ICRA 2022 and the Best Application Paper Finalists in IROS 2023. He won first place in the Robotic Grasping of Manipulation Competition-Picking in Clutter in ICRA 2024.

Chanqing Guo received the B.S. degree in Robotic Engineering from Shein-Ming Wu School of Intelligent Engineering, South China University of Technology, Guangzhou, China. He is currently pursuing the M.S. degree in Data Science and Information Technology at Tsinghua Shenzhen International Graduate School, Tsinghua University, Shenzhen, China.

His research interests focus on tactile perception, with particular applications in novel structure design and robot learning.

Junhao Gong received the B.S. degree in Electronic Information Engineering from Harbin Institute of Technology, Weihai, China, in 2024. He is currently pursuing his Ph.D. degree in Tsinghua Shenzhen International Graduate School, Tsinghua University, Shenzhen, China. His research interests mainly focus on tactile sensing and contact-aware robotic manipulation.

Chenxin Liang received the B.S. degree in Electronic Engineering from Tianjin University, Tianjin, China, in 2023. She is currently pursuing the Ph.D. degree with the Tsinghua Shenzhen International Graduate School, Tsinghua University, Shenzhen, China.

Her research interests include vision-based tactile sensing, 3D reconstruction, and depth estimation.

Wenhua Ding received the B.E. degree from the Shine-Ming Wu School of Intelligent Engineering, South China University of Technology, China, in 2024. He is currently pursuing the Master's degree at the Tsinghua Shenzhen International Graduate School, Tsinghua University, China.

His research interests include robotics and UAV-related sensing technologies.

Wenbo Ding received the BS and PhD degrees (Hons.) from Tsinghua University in 2011 and 2016, respectively. He worked as a postdoctoral research fellow at Georgia Tech from 2016 to 2019. He is now an associate professor and PhD supervisor at Tsinghua Shenzhen International Graduate School, Tsinghua University, where he leads the Smart Sensing and Robotics (SSR) group. His research interests are diverse and interdisciplinary, which include tactile sensing, embodied intelligence and AI4S.

He has received many prestigious awards, including the IROS New Generation Star, Gold Medal of the 47th International Exhibition of Inventions Geneva and the IEEE Scott Helt Memorial Award. He serves as the editorial board member and co-chair/area chair for several top journals/conferences.