

LightTact: A Visual–Tactile Fingertip Sensor for Deformation-Independent Contact Sensing

Changyi Lin^{*1}, Boda Huo^{*1}, Mingyang Yu¹, Emily Ruppel², Bingqing Chen², Jonathan Francis^{1,2}, Ding Zhao¹

<https://linchangyi1.github.io/LightTact>

Abstract—Contact often occurs without macroscopic surface deformation, such as during interaction with liquids, semi-liquids, or ultra-soft materials. However, most existing tactile sensors rely on deformation to infer contact, making such light-contact interactions difficult to perceive robustly. To address this, we present LightTact, a visual-tactile fingertip sensor that makes *contact directly visible* via a deformation-independent principle. LightTact features an ambient-blocking optical configuration that suppresses both external light and internal illumination at non-contact regions, while transmitting only the scattered light generated at true contacts. As a result, LightTact produces high-contrast raw images in which non-contact pixels remain near-black (mean gray value < 3) and contact pixels preserve the natural appearance of the contacting surface. Built on this, LightTact achieves accurate pixel-level contact segmentation that is robust to material properties, contact force, surface appearance, and environmental lighting. We further demonstrate that LightTact unlocks new robotic manipulation behaviors that require detection of extremely light contact, including water spreading, facial-cream dipping, and soft thin-film interaction. In addition, we show that LightTact’s spatially aligned visual-tactile images can be directly interpreted by vision-language models.

I. INTRODUCTION

Tactile sensing is essential for robotic manipulation, where robots must precisely regulate when, where, and how they physically interact with objects through contact [1, 2]. In recent years, vision-based tactile sensors (VBTs) have gained prominence due to their high spatial resolution and low cost [3, 4, 5, 6]. However, most VBTs sense contact *indirectly* by observing macroscopic deformation of a soft surface [7, 8, 9, 10, 11]. This deformation-dependent mechanism fundamentally requires measurable indentation to generate an informative tactile signal. In contrast, many real-world interactions involve *light* contacts that induce little or no surface deformation. Examples include interactions with liquids or semi-liquids, and gentle contacts with extremely soft or thin materials. Such scenarios are common in everyday tasks, such as spreading water on a tabletop, dipping facial cream, or detecting a sheet of food film, but they remain challenging for deformation-based tactile sensors.

To relax this constraint, several sensors attempt to bypass deformation by using frustrated total internal reflection (TIR) to highlight the contact regions of a transparent medium [12, 13, 14, 15]. However, ambient illumination and reflections from non-contact regions can also enter the camera, seriously

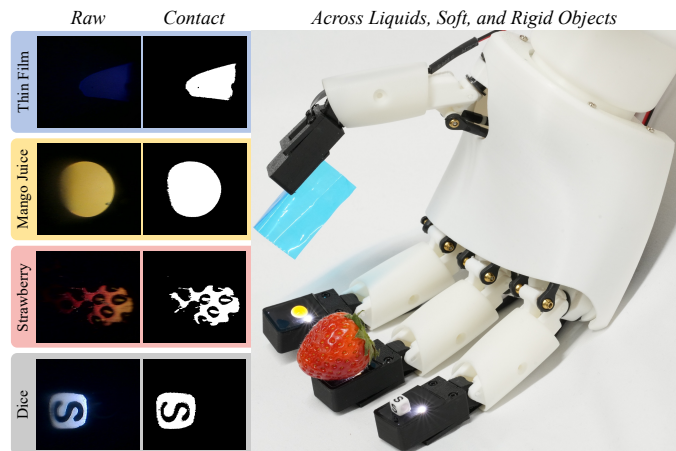


Figure 1. LightTact provides direct, pixel-level contact sensing across liquids, ultra-soft materials, and rigid objects, without requiring a minimum contact force. Its optical design produces high-contrast raw images in which non-contact pixels remain near-black, while contact pixels preserve the natural appearance of the contacting surface. LightTact is fingertip-sized and can be integrated into dexterous hands such as Amazing Hand [16].

corrupting the highlighted contact signal. As a result, TIR-based sensors typically function reliably only with monochromatic objects or under dark, tightly controlled environments—conditions incompatible with robotic tactile perception in everyday, unstructured settings. Therefore, designing a tactile sensor that can robustly detect contact *without* relying on surface deformation remains an open challenge.

In this work, we introduce LightTact, a camera-based fingertip sensor that makes physical contact *directly visible* in the raw image. LightTact enforces a bijective contact-visibility relationship: a pixel is visible *if and only if* it is in true physical contact. This enables robust pixel-level contact segmentation across liquids, ultra-soft materials, and rigid objects under diverse object appearances and bright ambient lighting (Fig. 1).

The key enabler is our *ambient-blocking* optical configuration that can, in theory, completely suppress both external light and internal illumination at non-contact regions, while transmitting only contact-generated scattered light to the camera through a transparent medium. We realize this behavior by adopting a side-view imaging optical layout with dedicated camera-LED-medium configurations, which jointly leverage refraction, specular reflection, and total internal reflection (TIR) to block non-contact light paths. To translate this principle into a practical robot fingertip, we co-design LightTact across optics, structure, materials, electronics, and fabrication. Finally, LightTact features the following properties:

* Equal contribution

¹ Carnegie Mellon University

² Bosch Center for Artificial Intelligence (BCAI)

- **High-contrast raw images** where non-contact pixels remain near-black (gray value < 3) while contact pixels preserve the surface appearance.
- **Pixel-level contact segmentation** with lightweight computation, robust to material properties, contact force, ambient lighting, and surface appearance.
- **Aligned visual–tactile signals** where both local appearance of the object and contact region of the sensor surface are naturally co-registered in the same RGB image.
- **Compact and compliant.** LightTact has a fingertip-scale size of $12 \times 18 \times 34.5$ mm and a soft surface, enabling direct integration into grippers and dexterous hands.
- **Affordable and reproducible.** The primary cost is the RGB camera; the remaining components can be fabricated with standard tools for under \$20.
- **Open-source.** We open-source both the hardware and software, together with a step-by-step fabrication tutorial to facilitate reproduction and further development.

Leveraging these properties, we integrate LightTact into a robotic system and demonstrate manipulation behaviors that remain difficult or infeasible for current advanced tactile sensors. In particular, LightTact enables reliable perception of extremely light contact during interaction with liquids, semi-liquids, and ultra-soft or thin materials, where deformation-based VBTs and force sensing provide uninformed or weak signals. Moreover, we demonstrate that LightTact’s visual–tactile outputs can be directly interpreted by Vision-Language Models (VLMs), enabling fine-grained reasoning such as resistor value inference for robotic sorting.

II. RELATED WORK

We review prior work in two aspects most relevant to this paper: (1) *robust contact sensing and segmentation* in vision-based tactile sensors (VBTs), and (2) *simultaneous visual–tactile sensing* for robotic manipulation.

A. Contact Sensing with Vision-Based Tactile Sensors

Vision-based tactile sensors (VBTs) use internal cameras to image tactile signals with high spatial sensitivity at low cost, but most detect contact *indirectly* via macroscopic deformation of a compliant surface. Common approaches include tracking painted markers [8, 10, 17, 18, 11, 19], observing magnified reflective patterns [20, 21], or measuring luminance variations of reflective layers [7, 9]. Because they require visible deformation, the surface must be soft, the contacting object relatively harder, and the applied force above a detectable threshold, limiting *light contacts* with negligible indentation (e.g., water, facial cream, thin films).

To decouple sensing from deformation, several sensors [12, 13, 14, 15] apply total internal reflection (TIR) within a transparent medium; contact locally frustrates TIR and increases reflected light, highlighting contact regions. However, this strategy is sensitive to ambient illumination: external light, reflections from non-contact portions of the object, and internal LED leakage can enter the camera through non-contact regions, corrupting the intended TIR signal. As a

result, these systems often require monochromatic objects or controlled dark environments.

In contrast, LightTact uses an ambient-blocking optical layout that suppresses both external and internal illumination at non-contact regions while capturing only contact-generated scattered light, keeping non-contact pixels near-black in raw images and enabling robust pixel-level contact segmentation across lighting conditions and object appearances.

Similar objectives appear in non-robotic optical devices, such as the Wyman–White fingerprint imager [22] and the Drawing Prism [23]. However, these systems target static capture or tracing and assume more relaxed constraints (e.g., controlled lighting and limited object types), often relying on bulky geometries, dark environments, or even a thin liquid layer to remove air gaps—conditions incompatible with dynamic robotic tactile perception in everyday settings. These conditions are incompatible with robotic tactile perception.

B. Visual-Tactile Multimodal Sensing

A prevalent strategy for combining visual and tactile modalities is to paint markers on a transparent gel [8, 24, 25, 26, 27, 28, 29]. However, this approach provides only sparse motion cues and inherently occludes visual information at marker locations. To preserve full-pixel visibility, alternative methods utilize translucent coatings paired with illumination switching [30, 31, 32], but they support only one modality at a time. While TIR-based designs [12, 13, 14, 15] can theoretically provide visual information, the bright internal illumination required for tactile sensing often causes over-exposure, degrading the external visual features. In contrast, LightTact achieves the simultaneous acquisition of robust pixel-level contact segmentation and natural visual appearance using a single camera, ensuring both modalities are spatially and temporally aligned by construction.

III. METHODOLOGY

Our goal is to design LightTact as a robotic tactile sensor capable of reliable, pixel-level contact segmentation across diverse objects and lighting conditions. To this end, we first introduce a novel optical layout and sensing principle and analyze its theoretical behavior in Section III-A. We then present how LightTact realizes this principle within a compliant, fingertip-scale form factor suitable for robotic integration in Section III-B. Following, we describe the key fabrication steps for constructing LightTact sensors in Section III-C. Finally, we introduce the procedures for sensor calibration and algorithms for contact segmentation in Section III-D.

A. Optical Layout and Sensing Principle

Figure 2 (a) shows the core optical elements of LightTact: a transparent medium that interacts with external objects, an internal LED for illumination, and a camera that observes light exiting the medium. Two surfaces of the medium play distinct optical roles: the *touching surface*, where contact occurs, and the *viewing surface*, through which light exits toward the camera. The remaining surfaces are coated matte-black to

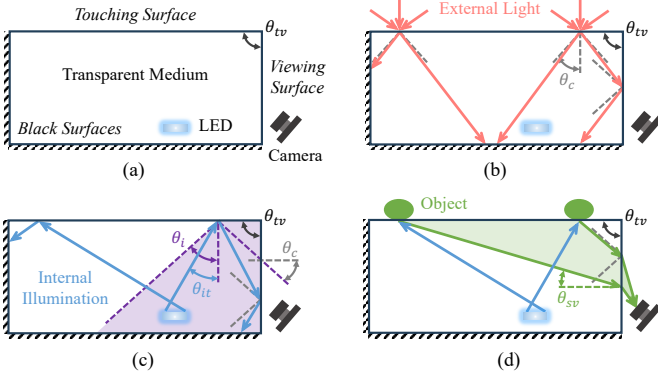


Figure 2. Optical layout and sensing principle of LightTact. (a) Core components: a transparent medium with a touching and viewing surface, an internal LED, and a camera. (b) External light entering through non-contact regions is rejected by refraction followed by total internal reflection (TIR). (c) Internal LED illumination reflecting from non-contact regions is also rejected by TIR when the LED is placed within the purple area to satisfy $\theta_{it} < \theta_i$. (d) At true contact, the air gap is removed and the contacting surface produces diffuse scattering. A subset of these rays reaches the viewing surface with incidence $< \theta_c$ and refracts toward the camera.

absorb any light that reaches them. At the touching surface, combining two light sources (external light and internal) with two region types (contact and non-contact) produces three relevant light behaviors for non-luminous objects: (1) external light incident on non-contact regions, (2) internal illumination reflected on non-contact regions, and (3) internal illumination scattered at contact regions.

Our objective is to obtain deformation-independent contact segmentation from the raw image. To achieve this, the optical layout must *completely suppress* external and internal light at non-contact regions, while *transmitting* only light generated at true contacts. The key is a side-view imaging layout: the camera observes the medium through a viewing surface that forms an angle θ_{tv} with the touching surface, rather than being parallel. This non-parallel geometry separates the light paths, so that only contact-generated light can transmit through the viewing surface and reach the camera. Table I summarizes these behaviors, and we detail them below.

a) External light rejection at non-contact regions: As illustrated in Fig. 2 (b), external rays approach the touching surface from a wide range of directions. The sensing region corresponds to the span between the two depicted ray bundles. When external light enters the medium through non-contact areas, it refracts at angles below the critical angle θ_c due to the air-medium index difference. A portion of these refracted rays then propagate toward the viewing surface. By choosing a sufficiently large wedge angle ($\theta_{tv} > 2\theta_c$), all such rays arrive at the viewing surface with incidence $> \theta_c$, causing them to undergo total internal reflection (TIR) and redirecting them toward the black surfaces. Consequently, no external light from non-contact regions reaches the camera.

b) Internal illumination rejection at non-contact regions: Internal LED illumination should contribute to the image only at true contacts. As shown in Fig. 2 (c), LED rays that reflect specularly off non-contact regions of the touching surface, must also reach the viewing surface with incidence $> \theta_c$ to

Table I
OPTICAL BEHAVIORS AND GEOMETRIC CONDITIONS FOR EACH LIGHT-SURFACE INTERACTION.

Source	Region	Optical behavior at surfaces		Design condition
		Touching surface	Viewing surface	
External	Non-con	Refraction	TIR	$\theta_{tv} > 2\theta_c$
Internal	Non-con	Specular reflection	TIR	$\forall \theta_{it} < \theta_{tv} - \theta_c$
Internal	Contact	Diffuse scattering	Refraction	$\theta_{tv} < \frac{\pi}{2} + \theta_c$

ensure TIR. This is guaranteed by constraining the LED position: the LED must lie entirely within the purple admissible region such that all emitted rays satisfy $\theta_{it} < \theta_i = \theta_{tv} - \theta_c$. Under this placement rule, every specularly reflected LED ray is rejected by TIR at the viewing surface.

c) Appearance capture at contact regions: When an object touches the medium, the air gap disappears and the contacting surface produces diffuse scattering of the LED illumination (Fig. 2(d)). This generates a wide angular spread of rays, which can reach the viewing surface with incidence $\theta_{sv} \in (\theta_{tv} - \frac{\pi}{2}, \frac{\pi}{2})$. To observe the contacting surface, a subset of these scattered rays must refract through the viewing surface and reach the camera. This requires that the incidence range $(\theta_{tv} - \frac{\pi}{2}, \frac{\pi}{2})$ intersects the transmissible range $(-\theta_c, \theta_c)$, which is satisfied when $\theta_{tv} < \frac{\pi}{2} + \theta_c$. Under this condition, a subset of diffuse-scattering rays transmits within the green regions and enters the camera, revealing the natural local appearance of the object.

Our optical layout enforces $\theta_{tv} > 2\theta_c$ so that any light entering through non-contact regions arrives at the viewing surface with incidence $> \theta_c$ and is therefore rejected by TIR, preventing it from reaching the camera. Strictly speaking, $\theta_{tv} > 2\theta_c$ is not the only feasible configuration: when $\theta_{tv} < 2\theta_c$, the camera could, in principle, be placed outside the refracted ray field from non-contact regions. In practice, however, this alternative sacrifices compactness and robustness, as it requires a larger camera offset and stricter suppression of stray reflections. We therefore adopt $\theta_{tv} > 2\theta_c$ as a simple and reliable default; a detailed analysis of the $\theta_{tv} < 2\theta_c$ case is provided in Appendix VI-A.

B. Sensor Design

As shown in Fig. 3(a), LightTact is extremely compact, with a fingertip-scale size of only 12mm $\times$ 18mm $\times$ 34.5mm. Figures 3(b) and 3(c) illustrate the internal assembly, which mirrors the optical system described in Section III-A: a transparent medium (2, 3), an internal LED (5), and a camera (7), all housed within a rigid shell (4), a surrounding black gel layer (1), and lightweight mounts (6, 8). The following paragraphs describe each component in detail.

Transparent gel and acrylic window. For most robotic applications, the touching surface should be compliant to support safe and stable interaction. Importantly, this compliance is a design preference rather than a sensing requirement, since our optical principle does not rely on macroscopic deformation. However, if the entire transparent medium is soft,

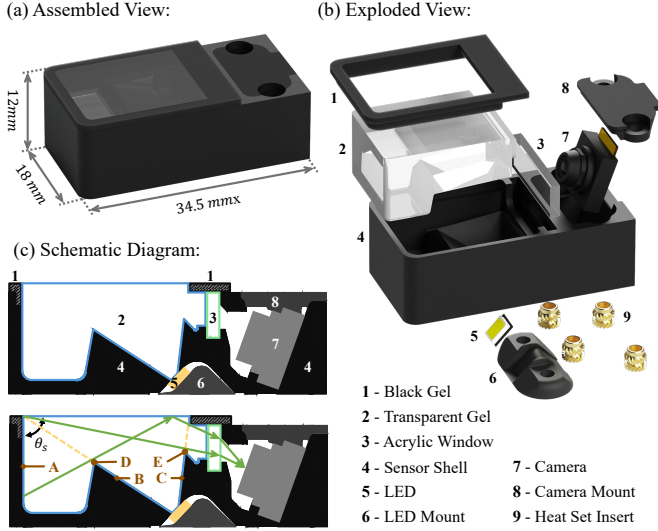


Figure 3. Design of LightTact. (a) Assembled view. (b) Components of LightTact illustrated in exploded view. (c) Schematic diagram.

the viewing surface would deform under load and introduce image distortion. To avoid this, we construct the medium as a composite: a soft, transparent gel that forms the touching surface, and a rigid acrylic window that defines a deformation-free viewing surface. The two materials have similar refractive indices (≈ 1.45), corresponding to a critical angle of $\theta_c \approx 43^\circ$. To satisfy the geometric conditions in Section III-A, namely $\theta_{tv} > 2\theta_c \approx 86^\circ$ to suppress non-contact light and $\theta_{tv} < 90^\circ + \theta_c \approx 133^\circ$ to transmit contact-scattered light, we set the touching and viewing surfaces to be perpendicular.

LED and camera. Internal illumination is provided by a single 2835-SMD LED, mounted at an oblique angle ($\approx 45^\circ$) to improve lighting uniformity across the touching surface. Imaging is performed by a compact UVC camera (OV5693) with a 120° field of view, placed close to the transparent medium to preserve a fingertip-scale form factor. Both components are secured with lightweight mounts that are attached to the sensor shell using M2 screws and heat-set inserts.

Sensor shell. The shell not only holds the optical components at their required poses, but also enforces three key geometric-optical constraints: the transparent gel’s geometry, the LED’s illumination envelope, and the camera’s effective viewing range. These design aspects are detailed below.

a) Transparent gel geometry: Although the shell is printed in black to approximate the ideal light-absorbing boundaries, practical materials still exhibit residual reflectance. As shown in Fig. 3(c), surface A can reflect stray light into the camera through non-contact regions. To mitigate this, we design surface A to be vertical ($\theta_s = \frac{\pi}{2}$), reducing the chance that specular reflections couple into the viewing path. For applications that require access to confined spaces, the shell can be made wedge-shaped. We analyze the case’s ($\theta_s < \frac{\pi}{2}$) optical consequences in Appendix VI-C.

b) LED illumination range: To prevent the LED from directly illuminating surface A, the shell includes two baffles (surfaces B and C) that bound the emission cone. Increasing

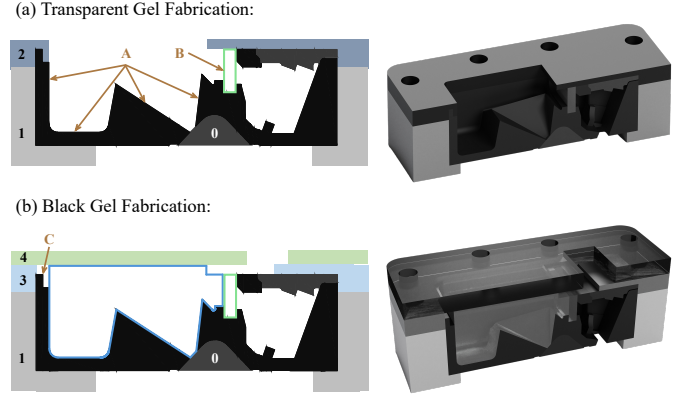


Figure 4. Gel fabrication for LightTact. (a) Casting of the transparent gel. The sensor shell, the mold 0, and the acrylic window together form a closed container. (b) Casting of the black gel. The top acrylic mold 4 ensures that the black gel cures to a uniform height around the transparent gel.

baffle height improves shielding but is limited by two factors: (1) if surface B extends above point D, the exposed inclined region becomes directly illuminated and increases non-contact light leakage; and (2) if surface C exceeds point E, it occludes part of the contact-generated scattered light from the left portion of the touching surface.

c) Camera’s effective viewing range: The shell also apertures the camera so that it observes only the touching surface. Allowing the camera to view outside this region would admit additional stray light (e.g., scattering from the rough top or bottom faces of the acrylic window) and interfere the pixels associated with the touching surface.

Black gel. To maintain compliance across the full contact area, we cast a soft black gel around the perimeter of the transparent gel, creating a smooth transition between the touching surface and the rigid shell. It also covers the top surface of the acrylic window, which is too rigid and rough to serve as part of the touching surface.

C. Sensor Fabrication

Most components of LightTact are straightforward to fabricate: the shell and mounts are 3D-printed with black PLA, the acrylic window is laser-cut from 1.5 mm acrylic sheet, the LED board is made by FPCB services, and the camera is off-the-shelf. The only nontrivial steps are casting the two gel layers, which require auxiliary molds and temporary inserts to define accurate internal boundaries (Fig. 4).

Transparent gel. We choose Silicones Inc[®] XP-565 silicone (cured hardness Shore A27) to make the transparent gel. The acrylic window is installed in the shell to form a sealed container for casting. To prevent silicone leakage into the LED compartment, we temporarily fill the LED recess with a solid insert (mold 0 in Fig. 4(a)). Two additional 3D-printed molds (labeled 1 and 2) are stacked around the shell to reserve the perimeter volume occupied by the future black gel.

Before pouring, we brush a high-absorption black pigment onto the relevant inner shell surfaces (region A in Fig. 4(a)) to better approximate the light-absorbing boundary condition. To

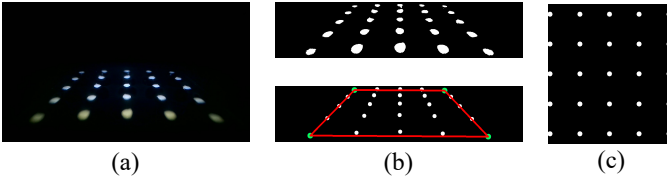


Figure 5. Camera calibration for LightTact. (a) LightTact captures imprints produced by a calibration tool with a 5×5 array of cylinders. (b) The imprints are segmented by thresholding and their center pixels are detected. (c) The final rectified and cropped image.

ensure strong bonding at the gel-acrylic interface, we apply a thin layer of silicone adhesive (Smooth-On® Sil-Poxy) to the acrylic face labeled B. We then degas and pour the XP-565 mixture until the cavity is fully filled, and cure it for ~ 4 hours on a heated 3D-printer bed at 50°C .

Black gel. The black gel is made from the same XP-565 silicone mixed with black pigment. After the transparent gel cures, we remove mold 2 and replace it with a second top mold (mold 3 in Fig. 4(b)) that defines the perimeter cavity for the black gel. Because this cavity is narrow and difficult to fill uniformly from above, we place an additional acrylic mold (labeled 4) on top of mold 3 and pour the mixture through the opening in mold 4, allowing it to flow uniformly around the perimeter. This “cap” mold also keeps firmly pressed against the touching surface, ensuring a consistent cured height. After curing under the same heated-bed conditions, we remove the molds and trim any excess gel around the fill opening.

As indicated by label C in Fig. 4(b), the shell includes a small raised perimeter step that creates the final three-layer edge profile (“shell – black gel – transparent gel”). This step is necessary because mold 4 slightly compresses the transparent gel during black-gel casting; after demolding, the transparent gel elastically recovers and can leave a thin lateral air gap beneath the black gel. Without the raised step, this gap would form a direct light path into the sensor. The raised perimeter therefore acts as a light-blocking barrier, preserving the dark-boundary assumption required by our optical design.

D. Contact Segmentation

Camera operation and sensor calibration. We disable auto-exposure and operate the camera at a fixed exposure time of 20 ms. This is crucial because auto-exposure would brighten LightTact’s predominantly dark background, amplifying sensor noise and potentially saturating bright contact regions. For robotic use, we further calibrate a mapping from image pixels to physical locations on the touching surface. Following 9DTact [9], we use a calibration tool with a 5×5 array of cylindrical bumps at 3 mm spacing. We press the tool onto the surface and capture the resulting imprint pattern (Fig. 5(a)). We segment the imprints via thresholding and extract their center pixels (Fig. 5(b)). We then use the four outermost corner imprints to estimate the average pixel spacing and define the target rectified grid. Finally, we compute a warping map from the detected centers to the rectified grid, and apply it to rectify and crop each raw image into a top-down view (Fig. 5(c)).

Segmentation algorithm. The optical design of LightTact

naturally produces a high-contrast representation of contact: non-contact regions remain near-black, while true contact results in localized brightness increases. Building on this property, our segmentation algorithm adopts a simple yet highly robust frame-differencing strategy. Before operation, LightTact records N no-contact frames and computes their pixel-wise average as a reference image I_{ref} . Given an incoming frame I , we compute a difference image $I_{\text{diff}} = I_{\text{raw}} - I_{\text{ref}}$, in which contact pixels appear as positive brightness changes.

To separate true contact from sensor noise, we apply a multi-condition brightness consistency test. A pixel at location (x, y) is classified as contact if any of the following holds: (1) the mean RGB increase exceeds t_0 ; (2) at least one channel exceeds t_1 ; (3) at least two channels exceed t_2 ; or (4) all three channels exceed t_3 . These complementary criteria capture both strong changes and weaker but consistent increases produced by contact. In our implementation, we use $N = 10$ and $t_0, t_1, t_2, t_3 = 25, 20, 30, 40$.

IV. EXPERIMENTS AND RESULTS

In this section, we conduct experiments to mainly answer the following questions: 1) Can LightTact achieve the intended light-suppression behavior and produce natural high-contrast contact images? 2) Can LightTact perform robust deformation-independent contact segmentation across diverse materials, appearances, contact forces, and ambient illumination? 3) What new manipulation capabilities does LightTact unlock when equipped on robots?

A. Light Suppression at Non-Contact Regions

We first evaluate whether LightTact achieves the intended light-suppression behavior under varying levels of internal illumination and external ambient brightness. We measure the brightness at the touching surface using a digital light meter. For each condition, we record (1) raw images without contact (non-contact images) and (2) raw images with a cylindrical object touching the sensing surface (contact images). We report the mean and standard deviation of the non-contact images in Table II, along with representative raw and segmented results.

Internal illumination. We first test LightTact in a dark environment and gradually increase the LED voltage. When the LED is off, both contact and non-contact images remain near-black as expected. As LED brightness increases to 430 Lux, the contact images begin to reveal clear visual appearance of the object while the non-contact regions remain extremely dark (mean gray value = 1.0). Further increasing the LED brightness (e.g., 1050 or 1670 Lux) causes only slight increases in non-contact intensity. However, these higher illumination levels overexpose the contact regions and degrade object appearance. To maintain near-black non-contact regions and capture clear object appearance, we set 430 Lux (corresponding to a 3.4 V driving voltage) as the default internal illumination for LightTact.

External illumination. Using the default LED setting, we then increase external illumination by gradually brightening an external lamp. As shown in Table II, LightTact maintains

Table II
INTENSITY STATISTICS AND REPRESENTATIVE IMAGES OF NON-CONTACT AND CONTACT CONDITIONS UNDER DIFFERENT INTERNAL AND EXTERNAL ILLUMINATION LEVELS.

Internal brightness (Lux)	0	430	1050	1670	430	430	430	430	430	430
External brightness (Lux)	3	3	3	3	740	1020	1510	2010	2540	3520
LightTact in environment										
Non-contact image mean	0.41	1.00	1.79	4.26	1.28	1.52	2.40	2.94	4.69	6.90
Non-contact image std	0.49	0.33	0.97	3.45	0.47	0.64	1.09	1.61	2.83	4.14
Non-contact raw image										
Contact raw image										
Contact segmentation										

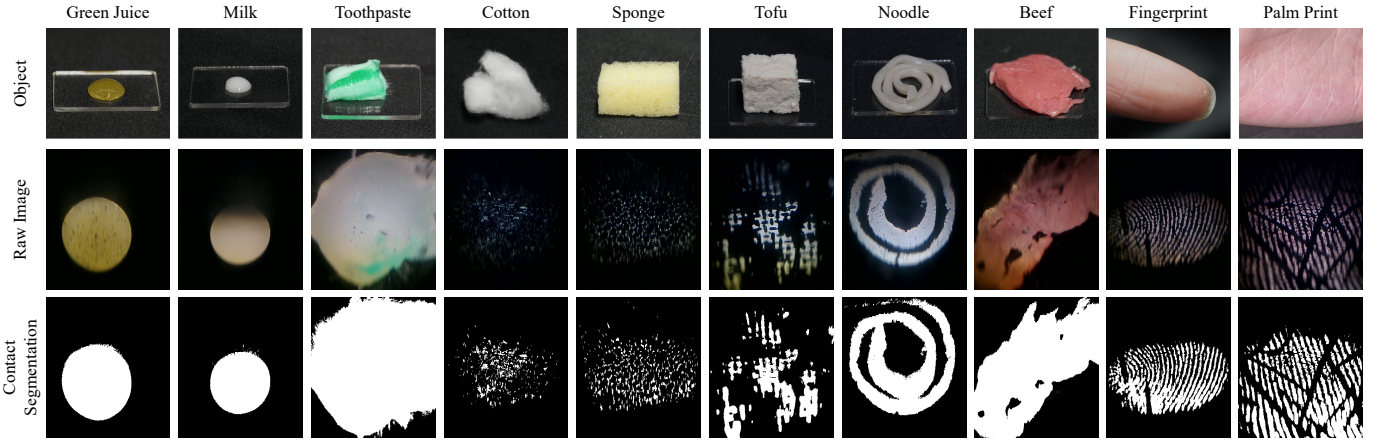


Figure 6. LightTact achieves robust contact segmentation with objects that make extremely light contact without producing macroscopic surface deformation, such as liquids, semi-liquids, and ultra-soft materials.

its non-contact pixels near-black (mean gray value < 3) even under strong ambient lighting up to 2010 Lux, which is far above typical indoor brightness (< 1000 Lux). At extremely bright illumination conditions (3520 Lux), the non-contact image becomes slightly brighter (mean gray value = 6.90) due to imperfect absorption on the internal surfaces of the sensor, as discussed in Section III-B. Nonetheless, contact regions remain clearly visible and segmentation remains robust.

These results demonstrate that LightTact consistently suppresses both external and internal light at non-contact regions while preserving natural appearance in contact regions. This behavior directly validates our optical design and highlights a key distinction from TIR-based sensors [12, 13, 14, 15], which fail to distinguish contact regions under bright environments.

B. Deformation-Independent Contact Segmentation

Extremely delicate contact sensing. Figure 6 evaluates LightTact’s ability to segment contact for liquids (green juice, milk), semi-liquids (toothpaste), and ultra-soft materials (cotton, sponge, tofu, noodle, beef, fingertip, palm) under bright

indoor illumination. Across all trials, objects are placed or gently touched onto the sensor without intentional pressing, resulting in minimal or no macroscopic surface deformation. Despite the absence of indentation cues, LightTact directly identifies the contact region from light intensity changes in the captured image, as its optical light-suppression mechanism produces clean, pixel-level segmentation even for visually complex materials. By contrast, deformation-based tactile sensors, including 9DTact [9], GelSight [7], and DelTact [11], do not yield reliable contact detection in this near-zero-force regime, since their measurements fundamentally rely on measurable deformation. Moreover, LightTact can detect contact under effectively zero applied force, as shown by the thin film hanging upside down on the downward-facing LightTact in Fig. 1; this further indicates that LightTact does not require a minimum force threshold to trigger contact perception.

Sensing across materials. Beyond liquids and soft materials, LightTact also provides robust visual–tactile sensing for rigid objects, as shown in Fig. 7 (textured joystick sur-

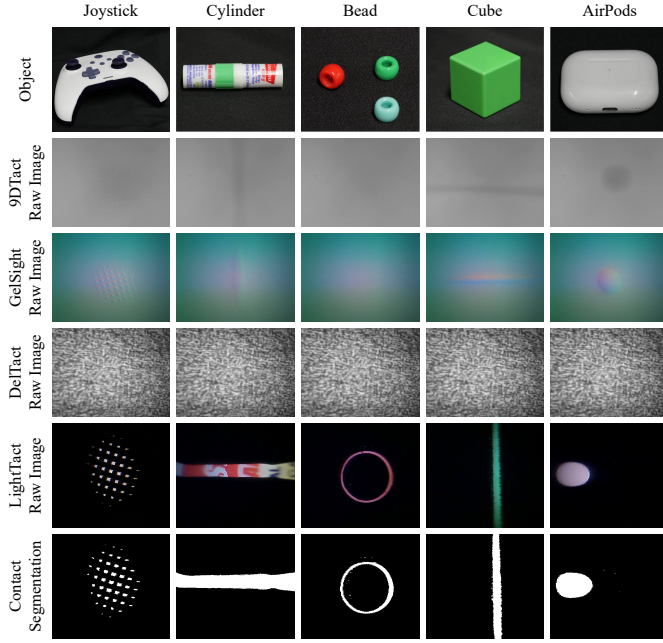


Figure 7. LightTact reliably senses both light and firm contact from rigid objects while preserving their visual appearance. In contrast, deformation-based tactile sensors fail under light contact, as the bead example illustrated.

face, complex-pattern cylinder, beads, cube, AirPods). Across both gentle placement and firmer pressing, LightTact consistently produces accurate contact segmentation while preserving natural visual appearance in the raw image. In contrast, deformation-based VBTSS become unreliable with rigid bodies when contact does not induce sufficient deformation (e.g., the bead case). Overall, these results show that LightTact performs accurate, pixel-level contact segmentation across liquids, semi-liquids, ultra-soft materials, and rigid objects. Such robust coverage across materials, appearances, and lighting is difficult to achieve with both TIR-based and deformation-based VBTSS.

C. Robotic Capabilities Unlocked by LightTact

To showcase the manipulation abilities enabled by LightTact’s direct, deformation-independent contact perception, we conduct a series of experiments on a UFACTORY xArm 7 robotic arm equipped with one or two LightTact modules mounted on its gripper. These demonstrations highlight two classes of capabilities:

- **Light contact sensing** for interaction with liquids, semi-liquids, and ultra-soft materials;
- **Multimodal reasoning** for fine-grained manipulation by directly prompting LightTact images to vision-language models (VLMs).

For all experiments, we additionally compare LightTact with a force-torque (FT) sensor and representative vision-based tactile sensors that are either open-source or commercially available, including 9DTact [9], GelSight-Mini [33], and DeiTact [11]. We refer to these as *baseline sensors* below.

1) *Light contact sensing of liquids, semi-liquids, and soft materials:* When a robot approaches an object for grasping or

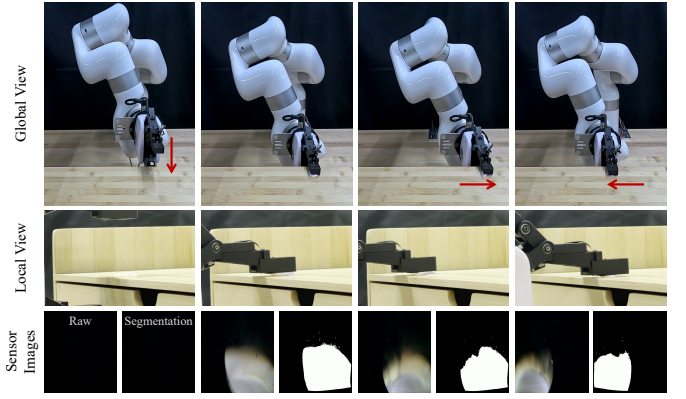


Figure 8. LightTact detects contact with water and segments the contact region, enabling the robot to spread water while maintaining contact and avoiding collision with the surface.

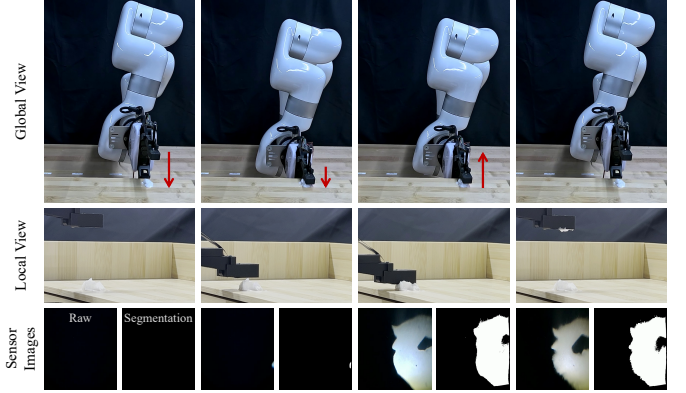


Figure 9. Using LightTact, the robot can dip facial cream using a contact-coverage criterion.

manipulation, the wrist-camera view is often occluded by the end-effector and the object, making tactile sensing essential for safe, closed-loop interaction. However, most VBTSS rely on contact-induced deformation and thus provide weak or unreliable signals when contact produces little-to-no indentation. In these regimes, LightTact remains effective: it detects contact at the gripper via a deformation-independent optical principle.

Liquid spreading. As shown in Fig. 8, a small amount of water is placed on a flat cabinet, and the robot orients LightTact downward. The gripper descends until contact is detected, then performs lateral sweeps while continuously adjusting its height to maintain contact with the water and avoid colliding with the surface. We regulate the end-effector height using a PD controller to maintain approximately 50% contact coverage. This closed-loop policy allows the robot to spread the water uniformly.

Semi-liquid dipping. As shown in Fig. 9, to collect facial cream, the robot moves its fingertip toward the cream, reduces its approach speed once contact is detected, and stops when more than 50% of the sensing surface is in contact. Notably, LightTact continues to detect the cream reliably during lifting, where essentially no pressure is applied on the sensing surface.

Extremely soft materials. As shown in Fig. 10, a human holds a small piece of food film that is ultra-soft, thin (0.1 mm), and light (0.05 g), and lightly touches it to a gripper

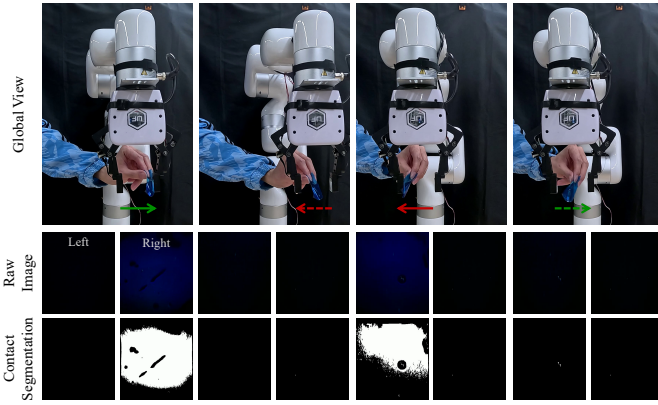


Figure 10. Two LightTact sensors enable responsive interaction with an ultra-thin film. Solid arrows indicate the end-effector motion commanded by contact on the corresponding sensor; dotted arrows indicate return to the centered position when no contact is detected.

equipped with two LightTact sensors. When the right sensor detects contact, the robot moves right; when the left sensor detects contact, it moves left; and when no contact is detected, it returns to a centered position. This behavior illustrates that LightTact can reliably sense minimal, delicate contacts and support responsive human-robot interaction.

Across all three tasks, the baseline sensors fail to produce usable signals because the contacts are too light to generate measurable forces or macroscopic deformation. These results underscore LightTact’s unique ability to enable robotic behaviors that require precise sensing of extremely light contact.

2) *Multimodal sensing and VLM-guided fine-grained manipulation*: Vision-language models (VLMs) are emerging as powerful general-purpose reasoning engines for robotics [34, 35]. However, conventional tactile inputs, either sparse force signals or deformation-based tactile images, provide limited appearance information and are often out-of-distribution for VLMs, making them difficult to interpret directly for manipulation [36, 37, 38]. In contrast, LightTact produces high-resolution, appearance-preserving images in which contact geometry and local object appearance are naturally aligned, making them directly compatible with VLM-based reasoning.

Resistor grasping and resistance reasoning. To demonstrate this capability, the robot is tasked with grasping and sorting resistors by their resistance values. The gripper closes until one LightTact detects more than 100 contact pixels, yielding a stable and gentle grasp. We crop the raw image around the contact region and prompt a VLM (GPT-5 Pro in our implementation) to infer the resistance value from the color bands. The robot then places the resistor into the corresponding cup. We evaluate this pipeline over 20 trials with 5 resistors and achieve 16 successful sorts (80% success rate). Most failures arise from the VLM confusing visually similar band colors (e.g., red vs. brown vs. orange).

In contrast, the baseline sensors do not provide interpretable appearance cues; therefore cannot support this VLM-guided sorting pipeline. We also compare against wrist-camera images: the VLM fails consistently because the resistor occupies a small fraction of the image and the background is cluttered.

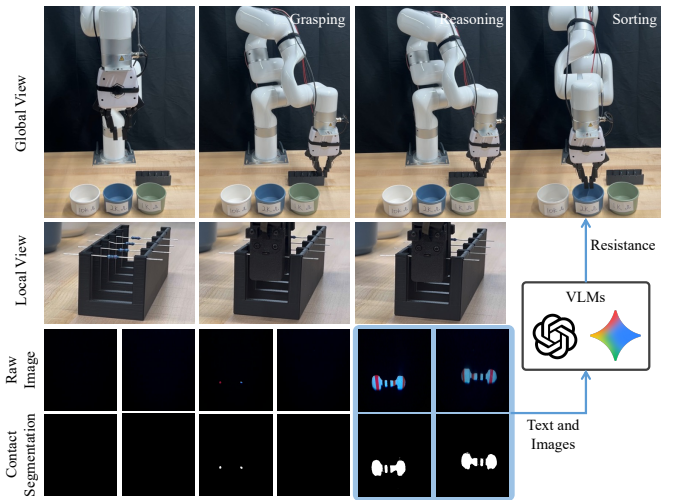


Figure 11. Beyond pixel-level contact sensing, LightTact preserves the natural appearance of the contacting surface, enabling VLMs to reason directly over its images. In this task, a VLM infers resistor values from color bands to guide robotic sorting.

Overall, these results demonstrate that LightTact can provide VLM-readable inputs without fine-tuning, enabling robots to obtain high-level task decisions directly through prompting.

V. ABLATION STUDY FOR SENSOR DESIGN

We evaluate ablation studies for several key design and fabrication choices introduced in Section III, with full results reported in Appendix VI-F.

- Black paint on the shell inner walls
- Acrylic mold for casting the black gel
- Edge indentation on the shell outer walls

VI. CONCLUSION

We presented LightTact, a compact visual-tactile fingertip sensor that enables deformation-independent, optics-based contact sensing. The core idea is an ambient-blocking optical configuration that suppresses both external light and internal illumination at non-contact regions, while transmitting only the scattered light generated at true contacts. This yields high-contrast raw images with near-black non-contact pixels and appearance-preserving contact pixels, making pixel-level contact segmentation simple and robust. Across systematic evaluations, LightTact maintains reliable contact segmentation over diverse materials (liquids, semi-liquids, ultra-soft materials, and rigid objects) and under strong ambient lighting. When integrated on a robotic arm, LightTact enables manipulation behaviors driven by extremely light contact, including water spreading, facial-cream dipping, and thin-film interaction, where force-based sensing and deformation-based VBTSs fail. This is promising for applications such as delicate handling of cosmetics, food, and biomedical samples, operating on wet surfaces, and enabling safe human-robot interaction. Moreover, LightTact provides spatially aligned visual-tactile images that can be directly prompted to vision-language models for fine-grained reasoning, demonstrated by resistor value inference for robotic sorting.

ACKNOWLEDGMENTS

We would like to thank Yuxiang Yang for his insightful discussions throughout the project.

REFERENCES

- [1] 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, 2009.
- [2] S. Luo, N. F. Lepora, W. Yuan, K. Althoefer, G. Cheng, and R. Dahiya, "Tactile robotics: An outlook," *IEEE Transactions on Robotics*, 2025.
- [3] H. Li, Y. Lin, C. Lu, M. Yang, E. Psomopoulou, and N. F. Lepora, "Classification of vision-based tactile sensors: A review," *IEEE Sensors Journal*, 2025.
- [4] S. Li, Z. Wang, C. Wu, X. Li, S. Luo, B. Fang, F. Sun, X.-P. Zhang, and W. Ding, "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.
- [5] Y.-H. Xin, K.-M. Hu, R.-J. Xiang, Y.-L. Gao, J.-F. Zhou, G. Meng, and W.-M. Zhang, "Vision-based tactile sensing: From performance parameters to device design," *Applied Physics Reviews*, vol. 12, no. 2, 2025.
- [6] K. He, "A survey of vision-based tactile sensors: Hardware, algorithm, application and future direction," *IEEE Transactions on Instrumentation and Measurement*, 2025.
- [7] 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.
- [8] A. Yamaguchi and C. G. Atkeson, "Combining finger vision and optical tactile sensing: Reducing and handling errors while cutting vegetables," in *2016 IEEE-RAS 16th International Conference on Humanoid Robots (Humanoids)*. IEEE, 2016, pp. 1045–1051.
- [9] C. Lin, H. Zhang, J. Xu, L. Wu, and H. Xu, "9dtact: A compact vision-based tactile sensor for accurate 3d shape reconstruction and generalizable 6d force estimation," *IEEE Robotics and Automation Letters*, vol. 9, no. 2, pp. 923–930, 2023.
- [10] B. Ward-Cherrier, N. Pestell, L. Cramphorn, B. Winstone, M. E. Giannaccini, J. Rossiter, and N. F. Lepora, "The tactip family: Soft optical tactile sensors with 3d-printed biomimetic morphologies," *Soft robotics*, vol. 5, no. 2, pp. 216–227, 2018.
- [11] G. Zhang, Y. Du, H. Yu, and M. Y. Wang, "Deltact: A vision-based tactile sensor using a dense color pattern," *IEEE Robotics and Automation Letters*, vol. 7, no. 4, pp. 10 778–10 785, 2022.
- [12] J. Y. Han, "Low-cost multi-touch sensing through frustrated total internal reflection," in *Proceedings of the 18th annual ACM symposium on User interface software and technology*, 2005, pp. 115–118.
- [13] K. Shimonomura, H. Nakashima, and K. Nozu, "Robotic grasp control with high-resolution combined tactile and proximity sensing," in *2016 IEEE International Conference on Robotics and automation (ICRA)*. IEEE, 2016, pp. 138–143.
- [14] S. Zhang, Y. Sun, J. Shan, Z. Chen, F. Sun, Y. Yang, and B. Fang, "Tirgel: A visuo-tactile sensor with total internal reflection mechanism for external observation and contact detection," *IEEE Robotics and Automation Letters*, vol. 8, no. 10, pp. 6307–6314, 2023.
- [15] Y. Sun, S. Zhang, Z. Chen, Z. Shen, F. Sun, C. Stefanini, D. Guo, S. Luo, J. Zhang, J. Shan *et al.*, "Soft contact simulation and manipulation learning of deformable objects with vision-based tactile sensor," *IEEE Transactions on Automation Science and Engineering*, 2025.
- [16] Pollen Robotics, "AmazingHand: Code and model to control the AH! (amazing hand)," <https://github.com/pollen-robotics/AmazingHand>, 2025, gitHub repository, accessed 2025-12-14.
- [17] A. Alspach, K. Hashimoto, N. Kuppuswamy, and R. Tedrake, "Soft-bubble: A highly compliant dense geometry tactile sensor for robot manipulation," in *2019 2nd IEEE International Conference on Soft Robotics (RoboSoft)*. IEEE, 2019, pp. 597–604.
- [18] S. Cui, R. Wang, J. Hu, J. Wei, S. Wang, and Z. Lou, "In-hand object localization using a novel high-resolution visuotactile sensor," *IEEE Transactions on Industrial Electronics*, vol. 69, no. 6, pp. 6015–6025, 2021.
- [19] L. Zhang, Y. Wang, and Y. Jiang, "Tac3d: A novel vision-based tactile sensor for measuring forces distribution and estimating friction coefficient distribution," *arXiv preprint arXiv:2202.06211*, 2022.
- [20] S. Saga, H. Kajimoto, and S. Tachi, "High-resolution tactile sensor using the deformation of a reflection image," *Sensor Review*, vol. 27, no. 1, pp. 35–42, 2007.
- [21] S. Saga, R. Taira, and K. Deguchi, "Precise shape reconstruction by active pattern in total-internal-reflection-based tactile sensor," *IEEE Transactions on Haptics*, vol. 7, no. 1, pp. 67–77, 2013.
- [22] W. Wyman, "Method for optical comparison of skin friction-ridge patterns," Aug. 17 1965, uS Patent 3,200,701.
- [23] R. Greene, "The drawing prism: a versatile graphic input device," in *Proceedings of the 12th annual conference on Computer graphics and interactive techniques*, 1985, pp. 103–110.
- [24] F. R. Hogan, J.-F. Tremblay, B. H. Baghi, M. Jenkin, K. Siddiqi, and G. Dudek, "Finger-sts: Combined proximity and tactile sensing for robotic manipulation," *IEEE Robotics and Automation Letters*, vol. 7, no. 4, pp. 10 865–10 872, 2022.
- [25] Q. Wang, Y. Du, and M. Y. Wang, "Spectac: A visual-tactile dual-modality sensor using uv illumination," in *2022 international conference on robotics and automation (ICRA)*. IEEE, 2022, pp. 10 844–10 850.
- [26] W. Fan, H. Li, W. Si, S. Luo, N. Lepora, and D. Zhang, "Vitactip: Design and verification of a novel biomimetic physical vision-tactile fusion sensor," in *2024 IEEE*

International Conference on Robotics and Automation (ICRA). IEEE, 2024, pp. 1056–1062.

- [27] L. Luo, B. Zhang, Z. Peng, Y. K. Cheung, G. Zhang, Z. Li, M. Y. Wang, and H. Yu, “Compdvision: Combining near-field 3d visual and tactile sensing using a compact compound-eye imaging system,” in *2024 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2024, pp. 262–268.
- [28] N. H. Nguyen, N. M. D. Le, Q. K. Luu, T. T. Nguyen, and V. A. Ho, “Vi2tap: A cross-polarization based mechanism for perception transition in tactile-proximity sensing with applications to soft grippers,” *IEEE Robotics and Automation Letters*, 2025.
- [29] Y. Li, Y. Chen, Z. Zhao, P. Li, T. Liu, S. Huang, and Y. Zhu, “Simultaneous tactile-visual perception for learning multimodal robot manipulation,” *arXiv preprint arXiv:2512.09851*, 2025.
- [30] F. R. Hogan, M. Jenkin, S. Rezaei-Shoshtari, Y. Girdhar, D. Meger, and G. Dudek, “Seeing through your skin: Recognizing objects with a novel visuotactile sensor,” in *Proceedings of the IEEE/CVF winter conference on applications of computer vision*, 2021, pp. 1218–1227.
- [31] E. Roberge, G. Fornes, and J.-P. Roberge, “Stereotac: A novel visuotactile sensor that combines tactile sensing with 3d vision,” *IEEE Robotics and Automation Letters*, vol. 8, no. 10, pp. 6291–6298, 2023.
- [32] S. Athar, G. Patel, Z. Xu, Q. Qiu, and Y. She, “Vistac toward a unified multimodal sensing finger for robotic manipulation,” *IEEE Sensors Journal*, vol. 23, no. 20, pp. 25 440–25 450, 2023.
- [33] GelSight. (2022) Gelsight-mini. [Online]. Available: <https://www.gelsight.com/gelsightmini/>
- [34] J. Gao, B. Sarkar, F. Xia, T. Xiao, J. Wu, B. Ichter, A. Majumdar, and D. Sadigh, “Physically grounded vision-language models for robotic manipulation,” in *2024 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2024, pp. 12 462–12 469.
- [35] G. R. Team, S. Abeyruwan, J. Ainslie, J.-B. Alayrac, M. G. Arenas, T. Armstrong, A. Balakrishna, R. Baruch, M. Bauza, M. Blokzijl *et al.*, “Gemini robotics: Bringing ai into the physical world,” *arXiv preprint arXiv:2503.20020*, 2025.
- [36] S. Yu, K. Lin, and H. Soh, “Demonstrating the octopi-1.5 visual-tactile-language model,” *arXiv preprint arXiv:2507.09985*, 2025.
- [37] J. Huang, S. Wang, F. Lin, Y. Hu, C. Wen, and Y. Gao, “Tactile-vla: Unlocking vision-language-action model’s physical knowledge for tactile generalization,” *arXiv preprint arXiv:2507.09160*, 2025.
- [38] Z. Cheng, Y. Zhang, W. Zhang, H. Li, K. Wang, L. Song, and H. Zhang, “Omnivtla: Vision-tactile-language-action model with semantic-aligned tactile sensing,” *arXiv preprint arXiv:2508.08706*, 2025.