R-Tac0: A Rounded High-Frequency Transferable Monochrome Vision-based Tactile Sensor for Shape Reconstruction

Wanlin Li[1*], Pei Lin[1,2*], Meng Wang[1], Chenxi Xiao[2],
Kaspar Althoefer[3], Yao Su[1†], Ziyuan Jiao[1†], Hangxin Liu[1†]

***Abstract*— Endowing the curved surfaces of rounded vision-based tactile fingers is essential for dexterous robotic manipulation, as they offer more sufficient contact with the environment. However, current rounded designs are constrained by a low sensing frequency (30–60 Hz) and the need for recalibration when adapting to new sensors due to the reliance on multi-channel captures, which hinders their performance in dynamic robotic tasks and large-scale deployment. In this work, we introduce R-Tac0, a low-cost rounded VBTS engineered for high-resolution and high-speed perception. The key innovation is a monochrome vision-based sensing principle: utilizing a black-and-white camera to capture the reflection properties of the compound rounded elastomer under monochromatic illumination. This single-channel imaging significantly reduces data volume and simplifies computational complexity, enabling 120 Hz tactile perception. A lightweight neural network can calibrate the sensor to achieve a depth reconstruction accuracy of 0.169 mm per pixel, while exhibiting surprisingly good transferability to new sensors. In experiments, we demonstrate the advantages of R-Tac0's rounded design by evaluating its performance under different contact angles, its high-frequency perception in slip detection, and its effectiveness in robotic dynamic pose estimation.**

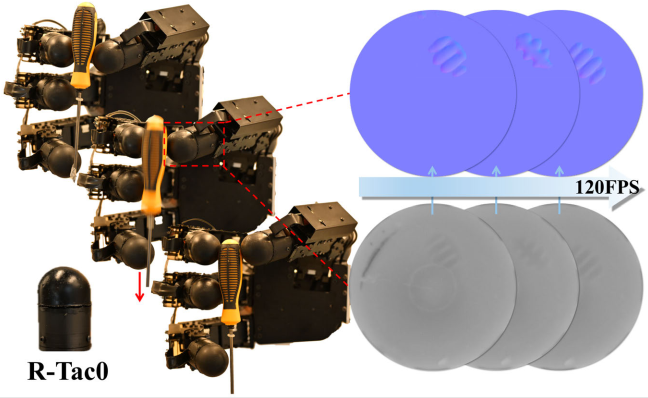


Fig. 1: **The R-Tac0 Sensor.** The bottom row shows a screw head rapidly sliding across the sensor surface. The middle row shows the image captured by the camera. The top row shows the reconstructed geometry.

I. INTRODUCTION

Tactile sensors are essential for robotic manipulation as they provide precise feedback on contact states, contact positions, and object surface characteristics, thereby allowing robots to perform more refined and accurate operations. In recent years, Vision-based Tactile Sensors (VBTSs) [1–9] have advanced significantly, with ever-higher perception resolution and ever-lower fabrication cost. Current VBTSs are primarily designed with a flat configuration, which limits their ability to adapt to the complex and varying contact states when deployed at the end-effector of robotic dexterous hands. Some research [10–14] has explored curved surface designs for integration with robotic hand fingertips. However, compared to non-vision-based tactile techniques [15] that operate above 100 Hz, VBTSs usually exhibit relatively low response frequencies, ranging from 30 to 60 Hz. This can lead to missed detections and motion blur during dynamic robotic tasks, significantly hindering their performances. The fundamental challenge lies in their dependence on multi-channel images to compute tactile information, which necessitates the use of RGB cameras and results in increased computational complexity and heating issues.

When deploying VBTSs on robotic multi-finger dexterous hand, factors such as calibration process and cost in large-scale deployments are also important considerations. The complexity of multi-channel data requires individual calibration for each newly manufactured sensor, and this calibration necessitate specialized equipment (*e.g.* CNC machines). These pose significant challenges to the efficient large-scale deployment of VBTSs at present.

Our response to this overall challenge is the development of R-Tac0, that leverages a *monochrome* vision-based sensing principle. The proposed sensor uses a low-cost black-and-white micro camera to capture the reflection properties of a coated semitransparent elastomer under white-lighting illumination, enabling high-frequency (120 Hz) tactile data computation through single-channel captures. We have also designed a calibration setup that can be manually printed and installed, providing convenience for researchers and practitioners to replicate. Additionally, due to the simplicity of the single-channel data, we have found that the depth calibration model based on differential images can be directly transferred to newly fabricated sensors. Therefore, R-Tac0 is a low-cost, monochrome vision-based tactile sensor capable of high-frequency and high-accuracy curved surface reconstruction. Its compact form factor and potential for cross-sensors transferability make it an ideal component

---

[*] Wanlin Li and Pei Lin contributed equally to this work.

[†] Corresponding authors.

[1] State Key Laboratory of General Artificial Intelligence, Beijing Institute for General Artificial Intelligence (BIGAI). Emails: {liwanlin, linpei, wangmeng, suyao, jiaoziyuan, liuhx}@bigai.ai.

[2] School of Information Science and Technology, ShanghaiTech University. Email: xiaochx@shanghaitech.edu.cn.

[3] Centre for Advanced Robotics @ Queen Mary (ARQ), Queen Mary University of London. Email: k.althoefer@qmul.ac.uk.

TABLE I: **Comparison of the proposed R-Tac0 with the state-of-the-art curved VBTSs**

Sensor	**Working Principle**	**Camera**	**Dimension** (mm)	**Cost** ($)	**Frequency** (Hz)	**Configuration**
TacTip [3]	Learning-based	Monocular RGB	$40 \times 40 \times 85$	-	90	Bionic fingertip
RainbowSight [16]	Photometric Stereo	Monocular RGB	$28 \times 28 \times 50$	-	30	Bionic fingertip
Omnitact [17]	Photometric Stereo	5 Cameras	$30 \times 30 \times 33$	3200	30	Bionic fingertip
GelTip [18]	Photometric Stereo	Monocular RGB	$30 \times 30 \times 100$	-	30	Bionic finger
InSight [19]	Photometric Stereo	Monocular RGB	$40 \times 40 \times 70$	-	40	Bionic finger
AllSight [20]	Photometric Stereo	Monocular RGB	$26 \times 28 \times 38$	30+	60	Bionic fingertip
DenseTact [12]	Learning-based	Monocular RGB	$32 \times 32 \times 43$	80-	30	Bionic fingertip
DIGIT Pinki [13]	Photometric Stereo	Monocular RGB	$15 \times 15 \times 15$	-	30	Bionic fingertip
GelStereo BioTip [21]	Binocular Stereo	Binocular RGB	$34 \times 28 \times 34$	-	60	Bionic fingertip
DTact [4]	Darkness Mapping	Monocular RGB	$32.5 \times 25.5 \times 25.5$	15	90	Non-planar
R-Tac0 (Ours)	Darkness Mapping	Monochrome	$30 \times 30 \times 43$	60	120	Bionic fingertip

for integration into standard robotic hands and multi-finger grippers. The key contributions of this work are outlined below:

- An adapted single monochrome camera with monochrome illumination on a multi-layer compound elastomer achieving high-speed, pixel-level sensing resolution.
- A novel curved sensor design and an accompanying streamlined, low-cost fabrication procedure.
- An easy-to-setup calibration method that leads to efficient and high-accuracy 3D shape reconstruction.
- An adapted neural network calibration model that can be effectively transferred to new sensors.
- Both hardware and code of R-Tac0 are open-sourced [1].

II. RELATED WORK

Current VBTSs tend to work on one of two possible solutions: marker array-based tracking or optical reflection-based surface reconstruction. **Marker array-based** VBTSs measure distributed multi-axis force/torque information and reconstruct surface shape by capturing the displacement of the marker array under the deformation of the contact module. TacTip [3], for example, uses a monocular RGB camera combined with learning-based techniques for sensor calibration, while GelStereo [21] uses binocular stereo vision for marker tracking. **Optical reflection-based** VBTSs reconstruct pixel-level high-resolution contact surface information by capturing variations in the light field of the contact module's surface. Both approaches have led to the development of compact VBTSs, that can be mounted and deployed on robotic dexterous hands.

One application of VBTS is to **estimate the pose of objects** during robotic grasping, manipulation, and interaction in dynamic and unstructured environments. Optical reflection-based VBTSs offer the advantage of higher spatial resolution in shape reconstruction, which results in more accurate positional information. Among these, RGB-based tactile sensors are now evolving from the original planar design [2] to novel curved surface designs [12, 14, 16–20], for improved performance in specific tasks or better alignment with the mechanical design of robotic end effectors. However, these sensors generally utilize monocular RGB cameras [2], binocular cameras [22], or depth cameras [23] to capture the deformation of the contact medium. Compared to tactile sensors based on capacitive [24–26], resistive [27–29], piezo-resistive [30–32] and magnetic [33–35] techniques, all of which can achieve a response frequency over hundred hertz (similar to robotic arms), VBTSs typically operate below 100 Hz. This lower operational frequency can result in missed detection and motion blur under dynamic contact and significantly limits their performance in dynamic tasks when integrated with robotic manipulators. Event-based cameras [36] offer an alternative way of capturing high-frequency images, but are usually prohibitively expensive (costing thousands of dollars).

As shown in Tab. I, R-Tac0 is designed in a bionic fingertip shape and measures high-resolution 3D geometry of curved surfaces using white light source and a monochrome black-and-white camera. Based on the principle of darkness mapping through optical reflection of the semitransparent elastomer, we proposed a novel camera-to-sensor calibration method for rounded surfaces that eliminates the need for specific equipment. R-Tac0 utilizes single-channel image calibration and reconstruction, which reduces data volume and enhances processing and transmission efficiency, resulting in exceptionally high-speed performance.

III. SENSOR DESIGN AND FABRICATION

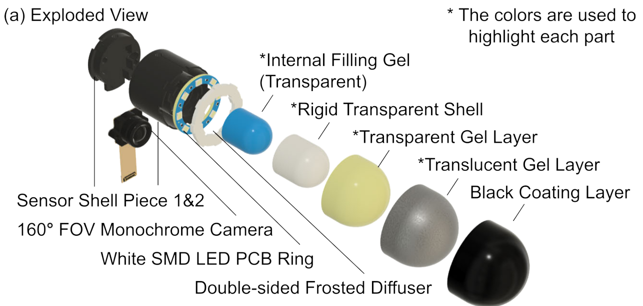
A. Design Criteria

R-Tac0 is a miniaturized vision-based tactile fingertip designed for integration into robotic end effectors and dexterous hands. The design criteria for the sensor are as follows: 1) **Compact Form Factor**: This allows the sensor to be installed in constrained spaces such as the fingertips of robotic hands; 2) **Round Shape**: The bionic fingertip design provides full-coverage tactile perception, to enable in-hand object pose estimation and manipulation; 3) **High-Resolution and High-Frequency Perception**: A high-speed camera has to capture distinct pixel-level surface deformations for precise shape reconstruction. 4) **Ease of Reproduction**: Sensor components should be readily accessible, easily fabricated, and cost-effective. The assembly process should be thoroughly

---

[1] https://github.com/bigai-ai/PP-Tac

(a) Exploded View



(b) Schematics

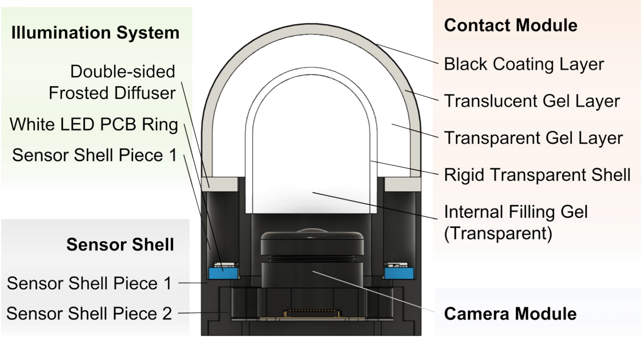


Fig. 2: **R-Tac0 Sensor Design.** (a) Exploded view of the sensor. (b) Schematics diagram of the four key components in the assembled R-Tac0 sensor.

described to facilitate robust reproduction by non-specialist developers.

B. *Sensor Fabrication*

As shown in Fig. 2, the R-Tac0 sensor is composed of four key components: the contact module, the illumination system, the camera, and the sensor shell.

1) Contact Module: To ensure clear and distinct contact deformation, this module is composed of five layers, arranged as follows, from the innermost to the outermost: internal filling gel, rigid transparent shell, transparent gel layer, translucent gel layer, and black coating layer.

Internal filling gel: This layer has two functions. Firstly, being transparent, it diffuses the light spot from the LED, as shown in Fig. 3, in which a clear distinction is evident between captures with and without the filling. Secondly, once cured, it acts as a glue, securing the rigid transparent shell to the sensor shell. The gel used is Smooth-on Solaris, mixed in a $1:1$ weight ratio (Part A and Part B), degassed in a vacuum chamber and cured at room temperature for approximately 24 hours.

Rigid transparent shell: This layer provides rigidity to the sensor structure during contact while allowing clear visibility of the elastomer deformation, and preventing delamination or tearing of the elastomer [16, 20]. Several methods can be used to manufacture this shell, including stereolithography (SLA), 3D printing with transparent materials [20] and casting with transparent epoxy resins [16]. In this study, off-the-shelf test tubes are utilized, which can be customized to the required thickness, radius, and height. Compared to those produced by 3D printing and casting techniques, glass tubes offer excellent clarity and an ultra-smooth surface, eliminating the need for post-processing. Prior to casting the soft transparent gel layer, the shell is primed with a clear silicone adhesive - we use Dow Corning PR1200 RTV.

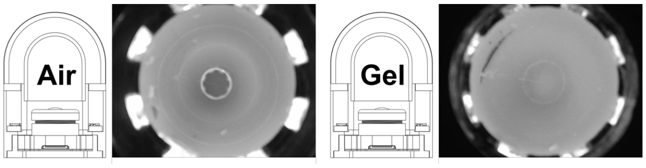


Fig. 3: **Effect of Internal Filling Gel.** Image captured by R-Tac0 sensor, with (right) and without (left) the internal gel layer. The layer helps eliminate the light source spots.

Transparent gel layer: This layer also serves two purposes. Firstly, it provides structural support between the translucent gel layer and the rigid transparent shell, as shown in Fig. 2. Secondly, it facilitates light diffusion, promoting more uniform illumination and mitigating potential reflection issues that can occur when light passes through air or other media [4]. We used stiff and durable PDMS - Dow Corning Sylgard 184 (Shore hardness 50A) - which offers both good optical transparency and resistance to tearing. The layer is fabricated through a two-piece molding, as shown in Fig. 4. The mold is created by installing the test tube into 3D-printed apparatus and pouring in Smooth-on MoldStar 30 - resulting in a smooth surface. A release agent (Mann Ease Release 200) is then sprayed inside the mold, and degassed PDMS (a $10:1$ mixture of Part A: Part B) is poured in, with the rigid shell press-fitted for casting. After curing for approximately 24 hours, the transparent gel layer, which bonds to the shell, is carefully removed from the mold.

Translucent gel layer and black coating layer: These two layers are designed to reflect internal light with varying brightness dependent upon contact. Taking into account the trade-off between fineness and depth in shape reconstruction [4], the thickness of the translucent gel layer is set to $1.5mm$ based on in-depth experimentation. It is worth noting that the translucent layer also determines the sensing range of the sensor. We use translucent silicone Smooth-on Ecoflex 00-30, which provides suitable hardness, excellent tensile strength. The black coating layer not only absorbs internal light and prevents interference from ambient light, but also serves as a wear-and-tear resistant protective layer with superior durability. To fabricate these layers, we use the same two-piece molding as described above. Firstly, a black mill-resistant mixture (Yonglihua Technology PTG-501) consisting of silicone oil, catalyst, and thinner in a $100:3:500$ ratio is prepared and sprayed onto the inner surface of the mold. The degassed silicone (a $1:1$ mixture of Part A: Part B) is then immediately poured in. The transparent gel layer is then press-fitted for casting, as shown in Fig. 4. After curing at room temperature for approximately 24 hours, the final bonded contact module is carefully removed from the mold.

2) Illumination System: Because of the optical properties of the contact module, a simple white LED ring is used to illuminate the sensor. Eight evenly spaced LUXEON 2835 4000K white SMD LEDs with corresponding 470

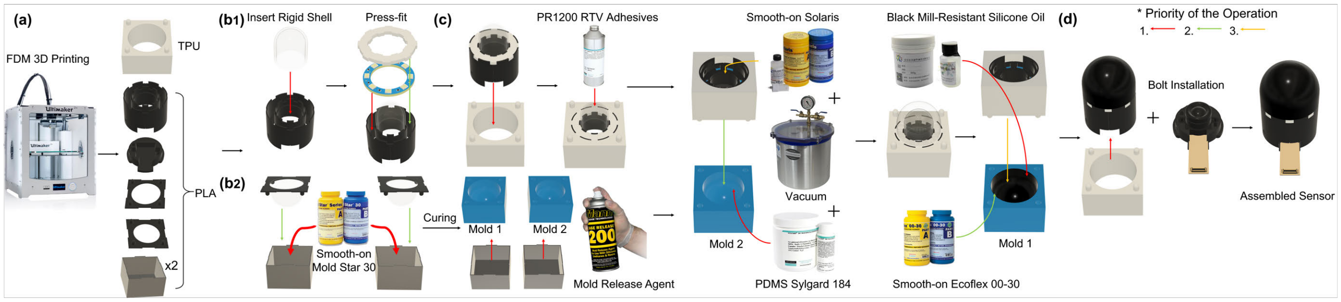


Fig. 4: **Fabrication Procedure of R-Tac0 Sensor.** Five main steps are depicted: (a) 3D printing the required parts. (b1) Fabricating and assembling the illumination system. (b2) Casting the molds. (c) Casting the contact module using two-piece molding technique. (d) Installing the camera and sensor shell.

Ohm resistors are soldered onto a customized annular PCB. The PCB is press-fitted into the sensor shell, and a laser-cut double-sided frosted diffuser is installed $10mm$ above the PCB to achieve effective light diffusion. The emitted light passes through the diffuser into the transparent and translucent layers to illuminate the black layer, which is then captured by the camera.

3) Camera: To achieve a compact design with a high response frequency, R-Tac0 employs a black-and-white CMOS OV9281 global shutter camera with a 160° wide field of view (FoV) lens. The camera works at a frame rate of 120 fps, with 640×480 resolution outputs. The camera module is secured to the sensor base with bolts and connected to the microcontroller, which is connected to the desktop via USB to stream single-channel MJPG format data. The exposure is set to manual mode to obtain consistent and stable output.

4) Sensor Shell: The two-piece black shell secures the camera, LED, diffuser, and contact module while also absorbing stray light, effectively guiding the lighting path, as shown in Fig. 2. We choose black PLA material to 3D print the sensor shell and connect the two pieces with small bolts.

Multiple R-Tac0 sensors are fabricated in this work. The overall dimensions of the sensors are $30mm \times 30mm \times 43mm$, and its weight is $35g$. The total cost of the sensor is around \$60, with the camera module (including the microcontroller) accounting for the majority of the expense at \$36. The contact module including molds is \$15, the LED ring with the diffuser is \$6, and the sensor shell is \$2. The entire fabrication process takes less than five days (most of which is taken up by silicone curing), which facilitates rapid reproduction and deployment in robotic applications.

IV. SENSOR CALIBRATION AND SHAPE RECONSTRUCTION

In this section, we describe procedures for (1) calibrating the camera, which is a prerequisite of sensor shape reconstruction; and (2) performing shape reconstruction based on camera model and image observations. For the camera's calibration, we aim to obtain parameters including the camera's intrinsic parameters ($\mathbf{M}$), distortion coefficients ($\mathbf{D}$), and extrinsic parameters ($\mathbf{R}$ and $\mathbf{T}$). As detailed in Sec. IV-A.

Next, we aim to accomplish 3D reconstruction based on the camera model obtained. One challenge associated with our sensor design is the curved sensor surface, which introduces varying curvatures across different regions. This incurs a generalization issue when relying solely on light intensity to infer surface geometry (such as using methods like linear regression based on color [4] or look-up tables [2]). To address this, we developed a robust neural network model, as detailed in Sec. IV-B.

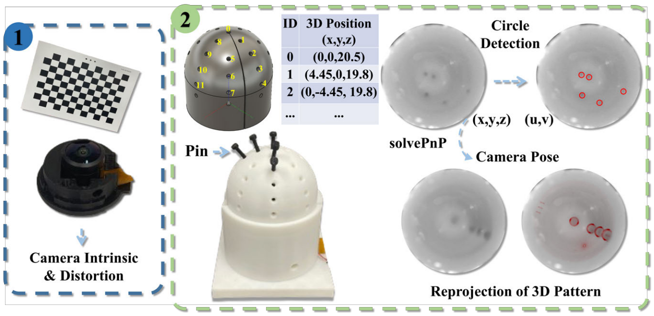


Fig. 5: **Camera Calibration.** We calibrate the intrinsic parameters and distortion using a calibration board. We 3D print a dome structure with predetermined holes, and by inserting pins at known coordinates, we are able to obtain pair-wise 3D and 2D points. The camera pose is then obtained using OpenCV's solvePnP function.

A. Camera Calibration

As a prerequisite for 3D reconstruction, the camera model needs to be fully determined. This involves: (1) obtaining the camera's intrinsic parameters M and distortion coefficients D using a checkerboard calibration method in OpenCV [37], and (2) determining the camera's position relative to the sensor surface (extrinsic parameters). Since it is intractable to attach a checkerboard pattern to the dome's inner surface, we propose a novel calibration approach that aims to obtain the sensor's extrinsic parameters. Specifically, we used a 3D-printed dome structure which is designed with holes at known locations. This allows pin insertions that make contact with the sensor at known spatial coordinates (x, y, z). Simultaneously, we recorded the pixel values (u, v) of each contact point. By using multiple pairs of 3D contact points and corresponding camera observations, the extrinsic parameters of the camera were calculated by solving the Perspective-n-Point problem.

$$(u, v) = MD[R \mid T](x, y, z) \tag{1}$$

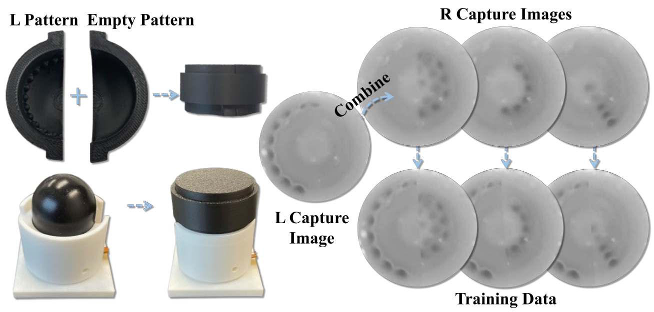


Fig. 6: **Groundtruth Capture. Above left:** We 3D-print semi-dome-shaped patterns with various internal protrusions. By pressing one protruded pattern and one empty pattern onto the R-Tac0, we completed capturing one data sample. Correspondingly, the ground truth depth is derived using the known camera parameters and the protrusions geometry. **Right:** By sampling and stitching together the images collected from different sides, we can quickly generate a large synthesized dataset using very few protruders.

Finally, we validated the accuracy of the obtained parameters by reprojecting the 3D contact points onto the camera image, which showed satisfactory alignment (See Fig. 5 bottom right).

B. *Sensor Calibration*

Prepare Training Data. Using the calibrated camera, we can generate depth maps relative to the camera coordinate by projecting arbitrary geometry, along with capturing corresponding grayscale images. However, collecting such pair-wise data individually is highly time-consuming. To address this, we have designed an efficient data collection method that significantly improves the process, achieving an exponential increase in data collection efficiency. The overall pipeline for this is shown in Fig. 6. Specifically, we 3D-printed a series of semi-dome-shaped components with internal protrusions. Each semi-dome can be combined with one empty semi-dome which has no protrusions inside. Since the geometry of these protrusions are known, we can directly establish the relationship between the camera's image observations and the corresponding ground-truth depth maps. After obtaining each pair of data, we rotate the combined two semi-domes by 180 degrees to generate another set of "mirrored" data from the flipped configuration. Benefiting from the use of grayscale images, this design enables a data augmentation technique where arbitrary pairs of left and right image observations can be stitched together, which significantly increases image diversity as shown in Fig. 6 (right). For example, with just 100 3D-printed patterns (semi-domes), this approach can yield $100 \times 100 = 10,000$ training samples, without using a CNC machine [12, 16, 38].

Depth Estimation Network. To take advantage of the camera's fast framerate, we propose a lightweight network structure to ensure high computational speed (see Fig. 7). The network's input is the differential image, calculated by subtracting a reference image (captured in a contact-free state) from the currently captured image. We then employed a two-layer Convolutional Neural Network (CNN), followed by a Multi-Layer Perceptron (MLP) to predict the differential depth for each pixel. The training can be formulized as

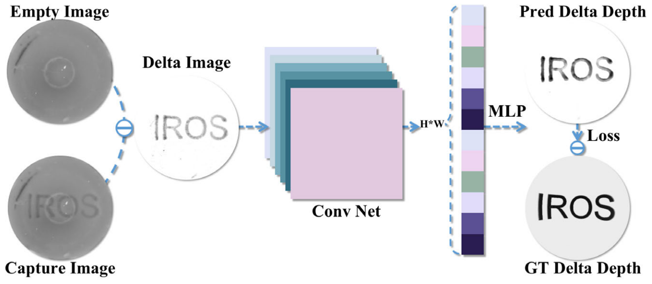


Fig. 7: **Pipeline of Depth Reconstruction.** To achieve high frequency and enhance the generalizability of our system, we employed a lightweight neural network for depth reconstruction. We input the differential image into a two-layer CNN, after which a two-layer MLP is utilized to identify the differential depth corresponding to each pixel. On an Nvidia RTX 4090, the computation time required for each frame is just $3.5ms$.

$$\Delta_{pred} = MLP(CNN(I_{capture} - I_{ref})), \tag{2}$$

$$L = \frac{1}{n}\sum_{i=1}^{n}(\Delta^{i}_{pred} - \Delta^{i}_{gt})^2, \tag{3}$$

where I denotes the captured gray-scale image, Δ denotes the differential depth map and n is the number of pixels. The network was trained by optimizing a Mean Squared Error (MSE) loss between the predicted differential depth and the ground truth differential depth. We obtained the complete surface depth map by combining the differential depth map with the sensor reference surface depth, which is projected using the camera parameters from Sec. IV-A based on the known sensor geometry.

Here we also provide the details of our proposed network architecture. The first CNN layer consists of 64 channels with a kernel size of 7, while the second layer comprises 128 channels with the same kernel size. The first layer of MLP reduces the feature dimension of each pixel to 64, and the second MLP layer further reduces the 64 dimension to 1.

C. *Generalizability to New Sensors*

For previous sensors with curved surfaces, the generalizability of learned depth estimators to new sensors remains an underexplored issue [12, 16, 38]. In this work, we address this generalizability challenge by exploring two solutions: (1) incorporating training data from multiple sensors, gathering 45,000 pairs of training data from five sensors; and (2) applying a fine-tuning technique after initial training. The results of these techniques are presented in Sec. V.

The depth reconstruction network was trained on an Nvidia RTX4090 for 10 epochs, taking around two hours. We also noticed the inference speed is ultra-fast. It takes only $3.5ms$ for each frame of image.

V. EXPERIMENTS

In this section, we evaluate the R-Tac0 sensor by performing a series of experiments. These include: (1) evaluating R-Tac0's performance in shape reconstruction quantitatively,

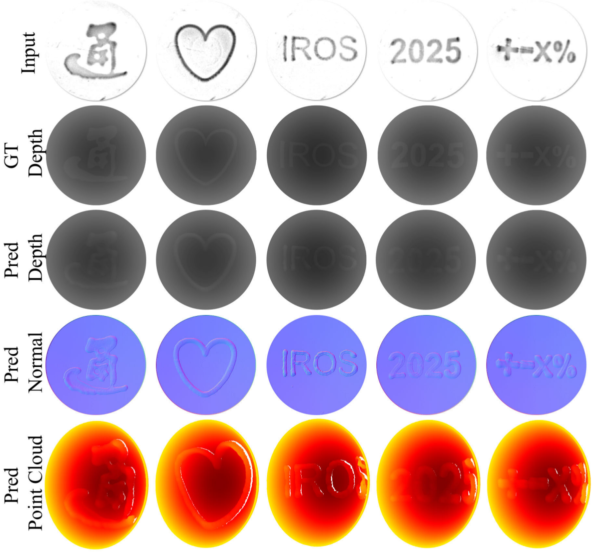


Fig. 8: **Reconstruction Results.** We showcased the prediction results of normals, depth, and point cloud using an unseen sensor.

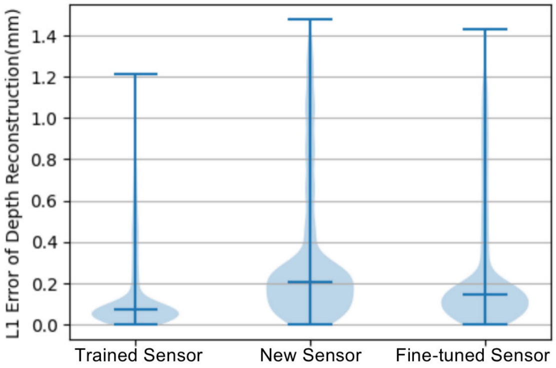


Fig. 9: **Quantitative Results of Depth Reconstruction.** We visualized the error distribution for all samples from the 3D reconstruction. The "Trained Sensor" refers to the sensor used for training data collection; the "New Sensor" is one that was not seen during training, and the "Fine-tuned New Sensor" incorporates the fine-tuning technique for an unseen sensor.

and qualitatively; (2) evaluating R-Tac0's high frequency on slip detection; (3) comparing R-Tac0 and a flat-shaped tactile sensor across varying contact conditions; (4) demonstrating R-Tac0's performance in dynamic pose tracking on its curved surface.

A. *Results of Depth Reconstruction*

First, we conducted experiments to demonstrate the performance of 3D reconstruction. Qualitative results are shown in Fig. 8, where various text patterns were pressed onto the sensor's surface. The results illustrate that the reconstruction performs well. It can be observed that all predicted depths, derived normals, as well as point clouds, have a clear correspondence with the input and the ground truth depth. Notably, all experiments were conducted using a new sensor not used during model training, highlighting the model's ability to generalize successfully to new sensors.

Next, we quantitatively analyzed the model's ability to generalize to new sensors excluded from training (Fig. 9). This was accomplished by employing test patterns with unseen, yet known geometries to indent the sensor surface. As a baseline, testing on a sensor included in the training set produced the best results (with a mean L1 error of $0.169mm$ and 75.78% of error samples below $0.1mm$). We then compared this baseline with a test conducted on a sensor unseen during training. This experiment resulted in a mean L1 error of $0.328mm$, with 47.76% of error samples below $0.1mm$. Finally, we fine-tuned the model using 100 data samples collected from this new sensor, spanning a total of 5 epochs. After fine-tuning, the mean L1 error decreased to $0.271mm$, with 71.32% of errors below $0.1mm$. This performance is comparable to that of the sensor used in training, demonstrating the effectiveness of the fine-tuning technique. It is noted that the maximum errors arise from position shifts of tiny surface defects introduced during the manufacturing process.

B. *Evaluation on High-Frequency: Slip Detection*

To assess the frequency performance of the sensor, a comparative analysis was performed to evaluate slip detection efficacy across multiple frame rates. As illustrated in Fig. 10, the R-Tac0 was mounted on a Franka Research 3 (FR3) robotic arm, and a weighted object was released from a fixed height to induce a consistent initial slip velocity on a ruler positioned at a predetermined location. The slip detection mechanism was implemented through a simplified algorithm: following the tactile sensor's initial contact establishment with the ruler, any detectable variations in the capture were classified as slip occurrences, thereby prompting the FR3 robotic arm to exert a controlled downward force towards the rule. Upon detection of slip, the magnitude of ruler displacement was quantitatively measured to assess the impact of different frame rates.

The sensor's frame rate was systematically configured to 120, 90, 60, and 30 fps, while the initial velocity of the ruler was maintained at 2 m/s through precise adjustment of the string length. For each frame rate, 50 experimental trials were conducted, yielding the following average ruler displacement measurements: $120\ fps: 3.1 \pm 1.2cm,\ 90\ fps: 3.8 \pm 1.3cm,\ 60\ fps: 5.1 \pm 2.3cm,\ 30\ fps: 8.2 \pm 3.2cm$. The experimental results demonstrate that higher frame rates significantly enhance the sensor's responsiveness to slip detection, enabling faster triggering of corrective force to mitigate slip. This effect becomes particularly pronounced at elevated slip velocities, where the system's rapid response capability is crucial. These experimental findings of slip detection underscore the importance of VBTS's frame rate in dynamic robotic tasks.

C. *Comparative Analysis of Curved and Flat Sensor Performance Across Varying Contact Angles*

To evaluate the advantage of the curved shape, a comparative experiment was performed to assess its performance with a flat tactile sensor [39] under varying contact angles. As illustrated in Fig. 11, both sensors were mounted on

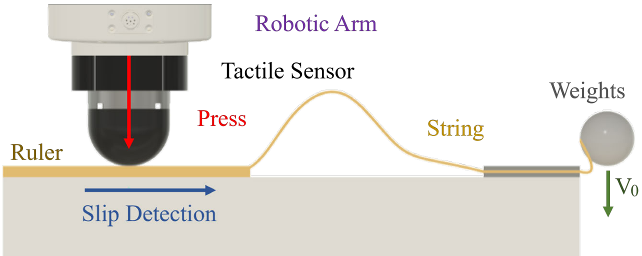
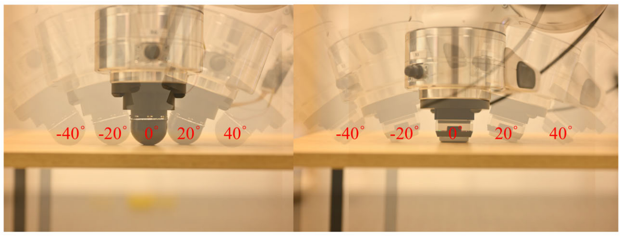


Fig. 10: **Experiment Setup for High-Frequency Evaluation.** For each trial, the ruler was positioned at the same initial location, and the robotic arm was controlled to ensure the tactile sensor to maintain a consistent position. A weighted object was released from rest at a predetertmined height, undergoing free-fall motion, while the velocity of the ruler was regulated by adjusting the length of the string. Upon detection of slip, the tactile sensor triggered the robotic arm to apply a controlled downward force to bring the ruler to a halt. By comparing the displacement of the ruler under different frame rates, the impact of sensor's frame rate on slip detection efficacy was evaluated.



(a) (b)

Fig. 11: **Varying Contact Conditions Applied to Curved and Flat Sensors.** (a) Pressing the R-Tac0 sensor against the tabletop at varying angles. (b) Pressing the flat tactile sensor [39] agiainst the tabletop at varying angles.

the end-effector of a Franka Emika FR3 robotic arm. The robotic arm was controlled to position the tactile sensors to make contact with the table at angles of -40°, -20°, 0°, 20°, and 40°, while maintaining a consistent contact depth for each interaction. For each angle, 20 trials were conducted, and the maximum depth readings from the sensor were recorded. The results demonstrate that the curved sensor consistently and accurately delivered stable readings across all contact angles. However, the flat sensor exhibited limitations in maintaining reliable readings in this case. When the flat sensor was in parallel contact with the table, it failed to produce repeatable measurements due to the deformation of its entire surface. Furthermore, at contact angles exceeding 40°, the flat sensor's limited contact area resulted in ineffective readings. These experimental findings underscore the superior adaptability of the curved sensor to diverse contact conditions.

D. Dynamic Pose Tracking on Curved Surface

Lastly, we demonstrate that the R-Tac0 sensor is highly suitable for dynamic robotic tasks requiring high resolution and precision. To showcase this, we applied Iterative Closest Point (ICP) [40, 41] to track the pose of a test sphere pressed on the sensor's surface. The tracking results are shown in Fig. 12. Using Open3D [42], we visualized the sphere's pose on the curved surface in real time, highlighting the curved sensor's high performance in dynamic pose estimation.

TABLE II: **Quantitative Results of Depth Readings from Curved and Flat Sensor Across Varying Contact Angles.** The robotic arm was controlled to position the sensor to make contact with the table at angles of -40°, -20°, 0°, 20°, and 40°, with a consistent contact depth of either 1 mm or 2 mm. For each combination of contact angle and depth, 20 trials were conducted, and the average and standard deviation of the maximum detected depth values from the sensor were calculated. For the flat sensor, no readings were obtained when the contact angle reached 40°, and these instances were recorded as N/A.

(mm)	-40°	-20°	0°	20°	40°
Curved + 1mm	1.13 ± 0.12	1.08 ± 0.10	1.07 ± 0.11	1.10 ± 0.12	1.09 ± 0.09
Flat + 1mm	N/A	1.41 ± 0.39	0.93 ± 0.32	1.36 ± 0.29	N/A
Curved + 2mm	2.21 ± 0.14	2.08 ± 0.10	2.14 ± 0.13	2.05 ± 0.07	2.12 ± 0.11
Flat + 2mm	N/A	3.26 ± 0.45	2.16 ± 0.31	3.02 ± 0.47	N/A

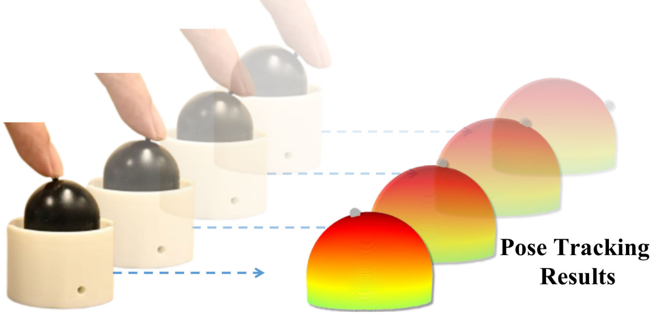


Fig. 12: **Dynamic Pose Tracking.** The left figure showcased pressing a small ball onto R-Tac0's elastomer. The right figure showcased the predicted ball location using ICP and reconstructed geometry.

VI. CONCLUSION & FUTURE WORK

This paper presents R-Tac0, a compact, low-cost vision-based tactile sensor that utilizes a single monochrome camera and a compound multilayer elastomer to achieve high-resolution and high-frequency curved surface reconstruction. To achieve this, we proposed an innovative calibration process that solely relies on 3D-printed setups, eliminating the need for additional costly machinery. R-Tac0 leverages a lightweight depth reconstruction network, achieving fast response times and robust generalization performance. Through a series of experiments, we demonstrate that the proposed sensor achieves accurate 3D reconstruction and can be transferred to new sensors. We validated the importance of high-frequency signals from the sensor in dynamic robotic tasks through slip detection. Additionally, we demonstrated the high adaptability of the curved shape under varying contact conditions through experiments involving contact at different angles. Finally, we showcased the application of the curved sensor in high-speed dynamic object pose tracking.

Future work will focus on deploying R-Tac0 sensors on robotic end-effectors, such as multi-finger grippers and hands, to perform various tasks. We will also explore solutions for deploying algorithms on robot onboard computers with limited resources, as well as exploring force perception.

Acknowledgement: We thank Mr. Mish Toszeghi from QMUL for proofreading.

REFERENCES

[1] S. Zhang, Z. Chen, Y. Gao, W. Wan, J. Shan, H. Xue, F. Sun, Y. Yang, and B. Fang, "Hardware technology of vision-based tactile sensor: A review," *IEEE Sensors Journal*, vol. 22, no. 22, pp. 21410–21427, 2022.

[2] 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.

[3] 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.

[4] C. Lin, Z. Lin, S. Wang, and H. Xu, "Dtact: A vision-based tactile sensor that measures high-resolution 3d geometry directly from darkness," in *IEEE International Conference on Robotics and Automation (ICRA)*, 2023.

[5] Y. Wu, Y. Chen, Z. Zhu, X. Qin, and C. Xiao, "Humanft: A human-like fingertip multimodal visuo-tactile sensor," *arXiv preprint arXiv:2410.10353*, 2024.

[6] W. Li, Z. Zhao, L. Cui, W. Zhang, H. Liu, L.-A. Li, and Y. Zhu, "Minitac: an ultra-compact 8 mm vision-based tactile sensor for enhanced palpation in robot-assisted minimally invasive surgery," *IEEE Robotics and Automation Letters*, 2024.

[7] W. Li, M. Wang, J. Li, Y. Su, D. K. Jha, X. Qian, K. Althoefer, and H. Liu, "L3 f-touch: A wireless gelsight with decoupled tactile and three-axis force sensing," *IEEE Robotics and Automation Letters (RA-L)*, vol. 8, no. 8, pp. 5148–5155, 2023.

[8] Z. Zhao, Y. Li, W. Li, Z. Qi, L. Ruan, Y. Zhu, and K. Althoefer, "Tac-man: Tactile-informed prior-free manipulation of articulated objects," *IEEE Transactions on Robotics*, 2024.

[9] Z. Zhao, W. Li, Y. Li, T. Liu, B. Li, M. Wang, K. Du, H. Liu, Y. Zhu, Q. Wang, *et al.*, "Embedding high-resolution touch across robotic hands enables adaptive human-like grasping," *Nature Machine Intelligence*, pp. 1–12, 2025.

[10] J. Hu, S. Cui, S. Wang, R. Wang, and Y. Wang, "Active shape reconstruction using a novel visuotactile palm sensor," *Biomimetic Intelligence and Robotics*, vol. 4, no. 3, p. 100167, 2024.

[11] M. H. Tippur and E. H. Adelson, "Gelsight360: An omnidirectional camera-based tactile sensor for dexterous robotic manipulation," in *IEEE International Conference on Soft Robotics (RoboSoft)*, 2023.

[12] W. K. Do and M. Kennedy, "Densetact: Optical tactile sensor for dense shape reconstruction," in *IEEE International Conference on Robotics and Automation (ICRA)*, 2022.

[13] J. Di, Z. Dugonjic, W. Fu, T. Wu, R. Mercado, K. Sawyer, V. R. Most, G. Kammerer, S. Speidel, R. E. Fan, *et al.*, "Using fiber optic bundles to miniaturize vision-based tactile sensors," *arXiv preprint arXiv:2403.05500*, 2024.

[14] B. Romero, F. Veiga, and E. Adelson, "Soft, round, high resolution tactile fingertip sensors for dexterous robotic manipulation," in *IEEE International Conference on Robotics and Automation (ICRA)*, 2020.

[15] M. L. Hammock, A. Chortos, B. C.-K. Tee, J. B.-H. Tok, and Z. Bao, "25th anniversary article: the evolution of electronic skin (e-skin): a brief history, design considerations, and recent progress," *Advanced materials*, vol. 25, no. 42, pp. 5997–6038, 2013.

[16] M. H. Tippur and E. H. Adelson, "Rainbowsight: A family of generalizable, curved, camera-based tactile sensors for shape reconstruction," in *IEEE International Conference on Robotics and Automation (ICRA)*, 2024.

[17] A. Padmanabha, F. Ebert, S. Tian, R. Calandra, C. Finn, and S. Levine, "Omnitact: A multi-directional high-resolution touch sensor," in *IEEE International Conference on Robotics and Automation (ICRA)*, 2020.

[18] D. F. Gomes, Z. Lin, and S. Luo, "Geltip: A finger-shaped optical tactile sensor for robotic manipulation," in *IEEE/RAS International Conference on Intelligent Robots and Systems (IROS)*, 2020.

[19] H. Sun, K. J. Kuchenbecker, and G. Martius, "A soft thumb-sized vision-based sensor with accurate all-round force perception," *Nature Machine Intelligence*, vol. 4, no. 2, pp. 135–145, 2022.

[20] O. Azulay, N. Curtis, R. Sokolovsky, G. Levitski, D. Slomovik, G. Lilling, and A. Sintov, "Allsight: A low-cost and high-resolution round tactile sensor with zero-shot learning capability," *IEEE Robotics and Automation Letters (RA-L)*, vol. 9, no. 1, pp. 483–490, 2023.

[21] S. Cui, S. Wang, C. Zhang, R. Wang, B. Zhang, S. Zhang, and Y. Wang, "Gelstereo biotip: Self-calibrating bionic fingertip visuotactile sensor for robotic manipulation," *IEEE/ASME Transactions on Mechatronics (TMECH)*, vol. 29, no. 4, pp. 2451–2462, 2024.

[22] 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. 25440–25450, 2023.

[23] N. Kuppuswamy, A. Alspach, A. Uttamchandani, S. Creasey, T. Ikeda, and R. Tedrake, "Soft-bubble grippers for robust and perceptive manipulation," in *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 9917–9924, IEEE, 2020.

[24] A. B. Dawood, H. Godaba, A. Ataka, and K. Althoefer, "Silicone-based capacitive e-skin for exteroception and proprioception," in *IEEE/RAS International Conference on Intelligent Robots and Systems (IROS)*, 2020.

[25] L. Du, "An overview of mobile capacitive touch technologies trends," *arXiv preprint arXiv:1612.08227*, 2016.

[26] Y. Zhang, X. Zhou, N. Zhang, J. Zhu, N. Bai, X. Hou, T. Sun, G. Li, L. Zhao, Y. Chen, *et al.*, "Ultrafast piezocapacitive soft pressure sensors with over 10 khz bandwidth via bonded microstructured interfaces," *Nature Communications*, vol. 15, no. 1, p. 3048, 2024.

[27] Z. Su, J. A. Fishel, T. Yamamoto, and G. E. Loeb, "Use of tactile feedback to control exploratory movements to characterize object compliance," *Frontiers in neurorobotics*, vol. 6, 2012.

[28] X. Hu, Z. Zhang, X. Chen, F. Wang, L. Shu, X. Xu, and X. Tao, "High-resolution, high-speed and low-cost flexible tactile sensor array system," *Measurement*, p. 115630, 2024.

[29] J. A. Fishel and G. E. Loeb, "Bayesian exploration for intelligent identification of textures," *Frontiers in neurorobotics*, vol. 6, p. 4, 2012.

[30] K. Motoo, F. Arai, and T. Fukuda, "Piezoelectric vibration-type tactile sensor using elasticity and viscosity change of structure," *IEEE Sensors Journal*, vol. 7, no. 7, pp. 1044–1051, 2007.

[31] W. Chen and X. Yan, "Progress in achieving high-performance piezoresistive and capacitive flexible pressure sensors: A review," *Journal of Materials Science & Technology*, vol. 43, pp. 175–188, 2020.

[32] J. Zhang, H. Yao, J. Mo, S. Chen, Y. Xie, S. Ma, R. Chen, T. Luo, W. Ling, L. Qin, *et al.*, "Finger-inspired rigid-soft hybrid tactile sensor with superior sensitivity at high frequency," *Nature communications*, vol. 13, no. 1, p. 5076, 2022.

[33] T. P. Tomo, A. Schmitz, W. K. Wong, H. Kristanto, S. Somlor, J. Hwang, L. Jamone, and S. Sugano, "Covering a robot fingertip with uskin: A soft electronic skin with distributed 3-axis force sensitive elements for robot hands," *IEEE Robotics and Automation Letters (RA-L)*, vol. 3, no. 1, pp. 124–131, 2017.

[34] Y. Yan, Z. Hu, Z. Yang, W. Yuan, C. Song, J. Pan, and Y. Shen, "Soft magnetic skin for super-resolution tactile sensing with force self-decoupling," *Science Robotics*, vol. 6, no. 51, p. eabc8801, 2021.

[35] L. Jamone, L. Natale, G. Metta, and G. Sandini, "Highly sensitive soft tactile sensors for an anthropomorphic robotic hand," *IEEE sensors Journal*, vol. 15, no. 8, pp. 4226–4233, 2015.

[36] 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 (T-RO)*, 2024.

[37] G. Bradski, "The OpenCV Library," *Dr. Dobb's Journal of Software Tools*, 2000.

[38] W. K. Do, B. Jurewicz, and M. Kennedy, "Densetact 2.0: Optical tactile sensor for shape and force reconstruction," in *IEEE International Conference on Robotics and Automation (ICRA)*, 2023.

[39] "Gelsight mini tactile sensor." `https://www.gelsight.com/gelsightmini/`.

[40] G. Izatt, G. Mirano, E. Adelson, and R. Tedrake, "Tracking objects with point clouds from vision and touch," in *2017 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 4000–4007, IEEE, 2017.

[41] S. Wang, Y. She, B. Romero, and E. Adelson, "Gelsight wedge: Measuring high-resolution 3d contact geometry with a compact robot finger," in *IEEE International Conference on Robotics and Automation (ICRA)*, 2021.

[42] Q.-Y. Zhou, J. Park, and V. Koltun, "Open3d: A modern library for 3d data processing," *arXiv preprint arXiv:1801.09847*, 2018.