

# Proximity and Visuotactile Point Cloud Fusion for Contact Patches in Extreme Deformation

Jessica Yin<sup>1\*</sup>, Paarth Shah<sup>2</sup>, Naveen Kuppuswamy<sup>2</sup>, Andrew Beaulieu<sup>2</sup>, Avinash Uttamchandani<sup>2</sup>, Alejandro Castro<sup>2</sup>, James Pikul<sup>1\*</sup>, and Russ Tedrake<sup>2\*</sup>

**Abstract**— Visuotactile sensors are a popular tactile sensing strategy due to high-fidelity estimates of local object geometry. However, existing algorithms for processing raw sensor inputs to useful intermediate signals such as contact patches struggle in high-deformation regimes. This is due to physical constraints imposed by sensor hardware and small-deformation assumptions used by mechanics-based models. In this work, we propose a fusion algorithm for proximity and visuotactile point clouds for contact patch segmentation, *entirely independent from membrane mechanics*. This algorithm exploits the synchronous, high spatial resolution proximity and visuotactile modalities enabled by an extremely deformable, *selectively transmissive soft membrane*, which uses visible light for visuotactile sensing and infrared light for proximity depth. We evaluate our contact patch algorithm in low (10%), medium (60%), and high (100%+) strain states. We compare our method against three baselines: proximity-only, tactile-only, and a first principles mechanics model. Our approach outperforms all baselines with an average RMSE under 2.8 mm of the contact patch geometry across all strain ranges. We demonstrate our contact patch algorithm in four applications: varied stiffness membranes, torque and shear-induced wrinkling, closed loop control, and pose estimation.

## I. INTRODUCTION

Humans are capable of an exceptionally diverse range of manipulation tasks, from precisely threading a needle to carrying another person. Tactile sensing has been studied with the goal of reproducing human-like manipulation abilities in robots. A popular tactile sensor paradigm are visuotactile sensors, which utilize high-resolution cameras to leverage advancements in computer vision techniques [1]–[6]. However, despite leveraging a soft surface as the contact interface, most visuotactile sensors rely on data processing techniques that lose functionality beyond small deformations, typically less than 1 mm. For example, dot tracking and optical flow algorithms used to estimate depth or track motion [7]–[10] in these sensors are not robust to either occlusions or the large marker displacements that occur during high deformation. This fundamentally restricts how much conformal contact can be made with the object.

The lack of conformal contact notably leads to small contact patches that are not innately informative for cm-scale objects without additional nontrivial implementations

<sup>1</sup>Department of Mechanical Engineering and Applied Mechanics and GRASP Lab at University of Pennsylvania, Philadelphia, PA, USA.

<sup>2</sup>Toyota Research Institute, Cambridge, MA, USA.

\*Corresponding authors: jessyin@seas.upenn.edu, pikul@seas.upenn.edu, russ.tedrake@tri.global

This work was supported in part by the National Science Foundation Graduate Research Fellowship Program under Grant No. 202095381 and by the National Science Foundation Emerging Frontiers in Research and Innovation (EFRI) award #1935294.

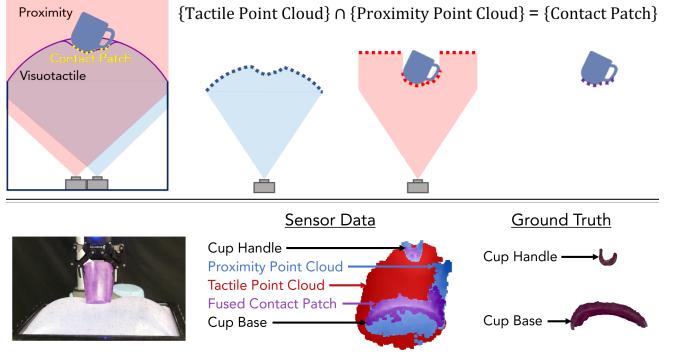


Fig. 1. We propose a mechanics-independent algorithm for estimating contact patches in challenging deformation regimes. We leverage a selectively transmissive soft membrane to provide simultaneous tactile and proximity point clouds. By fusing these two modalities through simply computing the intersection between the two point clouds, we obtain a high quality estimate of the contact patch. Because of our mechanics-independent approach, we avoid the non-linearities that many existing contact patches algorithm struggle with.

of tactile exploration policies [11], [12], data-driven pose estimation frameworks [13], [14], or integrating external vision sensors [15], [16]. While small contact patches may be acceptable for small, mm-scale, distinctly featured objects like screws and nuts, general real-world manipulation demands greater versatility and frequent interactions with larger objects. Limited conformal contact can also increase the difficulty of controlling the object. More compliance can be beneficial as it increases stiction and provides a larger margin of error for robots [17]. However, existing algorithms are poorly equipped to operate in the these regimes of large compliance.

Because deformation occurs beyond the contact patch, it is difficult to detect which parts of the membrane are in contact with the object exclusively from sensing the membrane’s deformation. The soft materials used to fabricate these compliant sensing surfaces have highly nonlinear mechanical properties that are difficult to model accurately and compute in real time for larger deformations.

In this work, we propose a contact patch segmentation algorithm that fuses proximity point clouds and tactile point clouds to detect contact patches of objects in extreme deformations of the sensing membrane. This algorithm leverages the multimodality of a large-scale proximity and visuotactile sensor enabled by a *selectively transmissive soft membrane*. The proximity depth modality is used to enhance tactile sensing functionality by providing essential object localization and geometry data during contact, independent from the mechanical behavior of the soft membrane. With this

key information and collocated proximity and tactile depth sensors, the intersection of the proximity and tactile point clouds provides the contact patch (Fig. 1). We make two key contributions:

- 1) We propose a novel contact patch segmentation algorithm that is *independent from mechanics of the deformable membrane*. This simple algorithm enables robustness to challenging deformation states, while still being fast enough for real-time control feedback. To the best of our knowledge, this is the first contact patch segmentation algorithm that fuses a point cloud proximity modality with tactile sensing.
- 2) We demonstrate applications of our contact patch algorithm in four use cases: varied stiffness membranes, torque and shear-induced out-of-plane membrane deformation, closed loop control for whole body manipulation, and pose estimation.

## II. RELATED WORK

Contact patch estimation is often the first step in processing tactile data for downstream tasks such as estimating object pose, contact force distributions and direction, and object classification. The three main approaches to contact patch estimation for soft tactile sensors with hyperelastic materials are to apply a threshold based on deformation, use data-driven techniques to classify contact patches, or solve a finite element method (FEM) model online.

Applying a threshold to membrane displacement is the simplest and fastest method to estimate contact patches [17]. A membrane displacement threshold number is manually chosen, and points over the threshold are considered as the contact patch. However, thresholds require manual tuning specific to the expected deformation and contact interface material, and do not generalize well across a large range of deformations.

Machine learning methods have been used to predict the contact patch, but they still have significant challenges: (1) the ability to generalize to a diverse set of object geometries, and (2) the laborious process of collecting real-world labelled datasets. Because the displacement to force mapping is dependent on object geometry and membrane mechanics, the datasets have to be very extensive to generalize to all possible objects that could be encountered by the sensor. Additionally, the ground truth for each image would have to be generated by a corresponding FEM simulation [18], [19], for which accuracy scales with computation cost.

With the FEM approach, each element of the elastomer is modelled as a hyperelastic material, for which the stress-strain relationship is derived from a strain energy density function [20]. Commonly used hyperelastic models for elastomers include the Neo-Hookean, Ogden, and Mooney-Rivlin formulations. However, to determine the stress-stretch curves necessary to use these models, samples of the material must be tested in uniaxial tension, pure shear, and equibiaxial tension [21]. Collecting this data requires specialized equipment and does not generalize well across manufacturing variabilities in different samples. Most significantly, FEM

approaches are computationally expensive and do not run in real-time. Linear approximations or assumptions of a constant stiffness matrix can be made to speed up runtime [22], [23], but these assumptions do not hold in cases of extreme deformations.

## III. SENSOR SYSTEM

The design and fabrication of the sensor builds upon [24]. The sensor provides synchronized 640x480 px depth maps in both the proximity and tactile modalities at 30 Hz.

The two internal cameras are the Intel Realsense L515 Time-of-Flight camera (proximity modality) and the Intel Realsense D405 (tactile modality). The IR light emitted by the L515 ToF camera travels through the membrane and does not sense the IR-transparent membrane itself. Although the membrane is somewhat translucent, the visual texture embedded in the membrane with random dye droplets enables the D405 to produce a point cloud of the membrane. The two cameras are placed immediately next to each other at the bottom center of the sensor. We design the active sensing region to be located where the FOVs of the D405's stereo depth and L515's ToF camera overlap; the entire exposed membrane is in view of both cameras.

The soft sensing surface is 355 mm x 205 mm and allows up to 85 mm of vertical displacement from a deflated membrane, before reaching the minimum sensing distances of the internal cameras. The sensor can be pressurized to inflate the membrane; the inflated height directly corresponds to how much the membrane can be vertically displaced. The proximity depth range extends approximately 100 mm beyond the membrane.

## IV. PROXIMITY AND VISUOTACTILE FUSION FOR CONTACT PATCH ESTIMATION

The fusion algorithm outputs a point cloud of the estimated contact patch, which is constructed from the intersection of the tactile and proximity point clouds (Fig. 1).

The data inputs to the algorithm are the proximity camera's RGB image, the proximity depth image, and the tactile depth image. The proximity camera's RGB modality, which observes the membrane, is only used to remove noise from the proximity depth data. The proximity depth and tactile depth maps are used to compute the contact patch.

1) *Frame Alignment*: First, the proximity RGB and tactile depth images are aligned to the proximity depth map. The proximity RGB and depth images are aligned using the Intel Realsense L515 intrinsics, since the depth and RGB cameras are located on the same device. The proximity depth and tactile depth images are aligned using a manually calibrated homography matrix.

2) *Proximity Pre-Processing*: Depending on the curvature of the inflated membrane, some areas of the membrane can be infrared reflective and produce noise in the proximity depth data. We convert the proximity RGB image to HSV and create a mask that removes any pixels that match the color of the membrane when no objects are in contact or near the membrane surface. This mask is tuned to environmental lighting conditions, but applies to all objects. The mask is

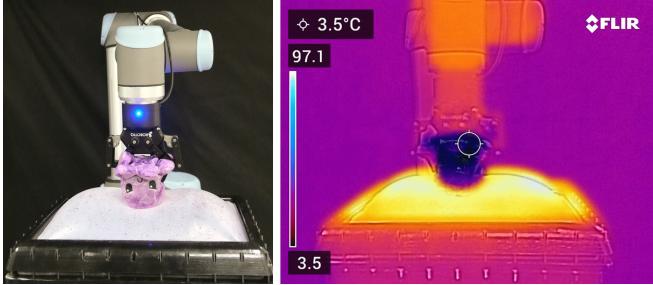


Fig. 2. The ground truth method for contact patch geometry exploits heat transfer between the object and sensor membrane at the contact interface. The object is painted with thermochromic paint that changes color when it comes into contact with the heated membrane.

then applied to the proximity depth image, which filters out the internal reflections of the emitted infrared light.

3) *Contact Patch Estimation*: The contact patch is estimated by computing the intersection of the tactile and proximity depth images. We compute an intersection mask with a pixel-wise comparison of distance values from the tactile and proximity depth maps:

$$|d_p - d_t| \leq t |d_p|$$

where  $d_p$  is the proximity distance,  $d_t$  is the tactile distance, and  $t$  is the tolerance. If the difference in tactile and proximity distance values at a pixel satisfy the inequality, the pixel is considered a match and part of the contact patch. We manually tune the value of  $t$  to produce the contact patch that best matches ground truth, and we found that  $t = 0.03$  works well; the tolerance distance values range from 6mm-8mm. The intersection mask is applied to the proximity depth image to produce the contact patch estimation. We use the proximity depth image instead of the tactile depth image because it is significantly more robust to inconsistent room lighting conditions. The contact patch depth image is then projected to a point cloud and we apply a statistical outlier rejection filter to remove any extraneous points before outputting the final contact patch point cloud estimation. Although this step does require some manual tuning of  $t$ , it generalizes across all strain ranges and objects.

## V. EXPERIMENTAL METHODS

### A. Ground Truth

Our ground truth method leverages thermal conduction at the contact interface, similar to [25], [26] (Fig. 2).

We paint the object dataset with three coats of thermochromic paint (Elmers), which changes color from purple (cold) to pink (hot) with temperature. The objects are cooled to  $3.5^{\circ}\text{C}$  until the purple color is opaque. The membrane is then heated to  $70^{\circ}\text{C}$ - $100^{\circ}\text{C}$  with a heat gun. The precise

Object	Membrane Temperature ( $^{\circ}\text{C}$ )	Contact Time (s)
octopus	70	5.5
cup	100	9
cube	95	9

TABLE I

Parameters for ground truth method.

temperature of the membrane and contact time is tuned per object to prevent noise from thermal radiation; the chosen temperature prevents color change when the object is within 1 mm of the membrane, but enables color change when contact occurs. The temperatures and contact time vary per object because each object has a different specific heat capacity (Table I). We monitor the spatial uniformity of the membrane and object temperature with a FLIR thermal camera.

The thermochromic paint changes color where contact occurs. We scan the object with a high-resolution 3D scanner (Space Spider, Artec) and threshold the scan by color to produce the ground-truth contact patch. This method captures the maximum contact patch, which we assume to occur when the membrane is at its highest deformation during its interaction with the object.

### B. Membrane Strain Measurement

We use the following method to correlate displacement of the object into the membrane to 2D strain in the membrane (Figure 3). First, we use a laser-cut stencil to pattern an IR and visibly opaque membrane with a uniform grid of dots with known spacing (10mm). This membrane is fabricated to have the same thickness and physical dimensions as our sensing membrane. We use the Realsense L515 to capture an RGB-D point cloud of the patterned membrane as the robot arm applies a vertical displacement with an object from our object dataset. We measure the distances between the centroids of each dot, which we use to calculate the maximum in-plane strain along the length and width of the membrane. The calculated maximum strain is used to find the vertical displacement required to impart the desired strain for our experiments: low (10%), medium (60%), high (100%).

### C. Baselines

For the tactile-only and membrane mechanics model baselines, we use an unpatterned IR-opaque silicone membrane with the same physical dimensions as the selectively transmissive membrane. We use the Realsense L515 ToF depth

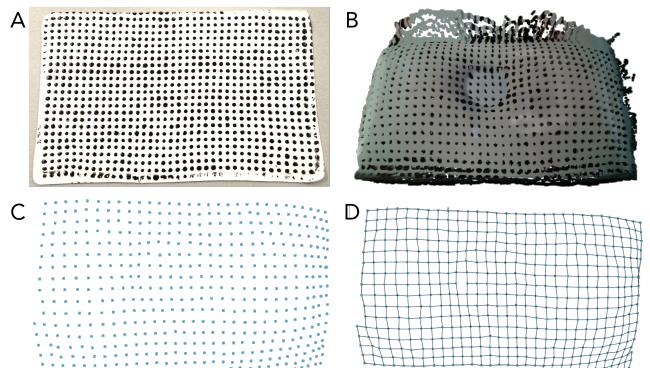


Fig. 3. Method used to calculate membrane strain. A) Uniform grid of dots with known spacing is patterned on the membrane. B) Object (octopus, not shown) is pressed into the membrane and the RGB-D point cloud is captured. C) The centroids of the dots from the RGB-D point cloud are isolated. D) The distances between the centroids are measured to calculate strain.

Algorithm	Computation Time (s)
Fusion (Ours)	0.07
Proximity-Only	0.03
Tactile-Only	0.03
Mechanics Model	0.3

TABLE II

Time to compute one frame using an Intel Core i7-8650U CPU.

camera to collect data for both of these baselines, since using a ToF depth camera with this sensor system can replicate the Soft Bubble sensor [5] and fulfill the input requirement of ToF depth data for the membrane mechanics model. For the proximity-only baseline, we use the selectively transmissive membrane and depth data from the Realsense L515 ToF camera. The computation times for each approach are in Table II.

1) *Tactile-Only*: The tactile-only approach exclusively uses a tactile point cloud and a thresholding scheme to estimate the contact patch. In this work, the threshold is implemented as follows: the contact patch is identified as the highest 60% of deformed points on the membrane, compared to a reference point cloud of an inflated membrane with no object-induced deformation. We found that an absolute threshold approach (i.e., points below a certain distance), although common for less deformable visuotactile sensors, gives much worse contact patch estimations and would not be a realistic baseline.

2) *Proximity-Only*: The proximity-only approach exclusively uses the proximity point cloud. This baseline uses a reference point cloud of the inflated membrane with no object-induced deformation. The contact patch is estimated as all points closer to the camera than the reference point cloud.

3) *Membrane Mechanics Model*: We use the first principles continuum mechanics model presented in [23] as a baseline. The model predicts the linear elastic deformation of a mesh of the membrane given the pose and geometry of a rigid object in contact with the sensor. Then, the model is used to solve the inverse problem of estimating the contact patch based on tactile depth and internal air pressure data from the sensor, using a sparse convex Quadratic Program formulation for real-time solutions. The model assumes that the contact surface is frictionless, deforms linearly, and does not wrinkle or fold.

## VI. EVALUATION

### A. Experiment Setup

We measure the accuracy of the estimated contact patch when the membrane is in low (approx. 10%), medium (approx. 60%), and high (approx. 100%+) strain states. We use a dataset of three objects: a stuffed octopus, a Rubiks cube, and a cup. These objects were chosen to represent a diverse set of geometries: edges (Rubiks cube), curves (octopus), and non-convex, separate contacts (cup). Each object is held by a Robotiq gripper mounted to a UR10 robot arm and pushed into the sensor. The vertical displacement of the object in each trial induces the desired strain in the membrane: 4 mm for low strain, 24 mm for medium strain, and 120 mm for high strain. For high strain experiments, the

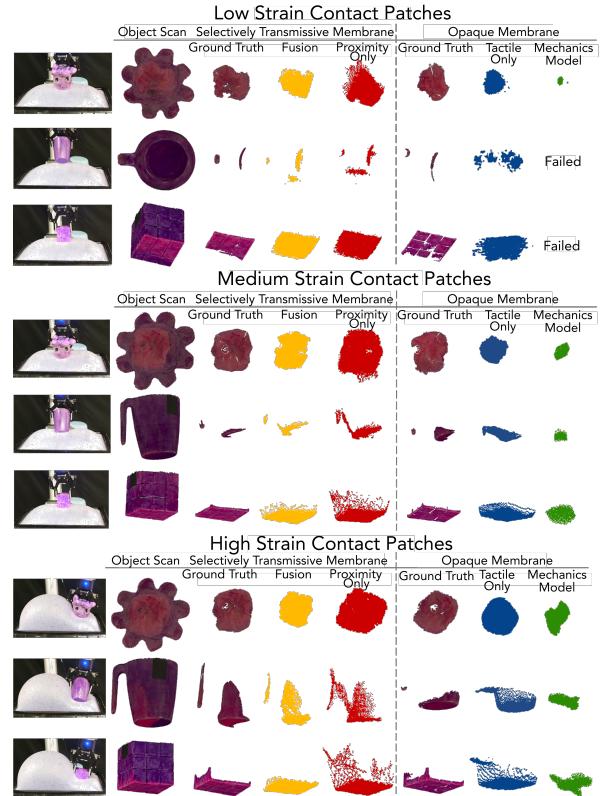


Fig. 4. Visualizations of the estimated contact patch across low, medium, and high strain states of the membrane. Our proposed algorithm, Fusion, fuses the visuotactile and proximity modalities of the sensor enabled by the selectively transmissive membrane. The baselines for comparison are Proximity Only, Tactile Only, and Mechanics Model. The ground truth is a color-segmented 3D scan of the object, where pink designates contact.

object is pressed off-center to allow the uncontacted region to expand outwards and produce high strains in the membrane.

We compare our results to ground truth and three baselines: tactile-only, proximity-only, and the mechanics model-based method. The metric for geometric accuracy of the contact patch point clouds is the *average symmetric root mean squared error* (RMSE) of the estimated contact patch point cloud and the ground truth point cloud after manual alignment. For the membrane mechanics model-based method, we crop the estimated contact patch to exclude the outer edges of the sensor interface since the model cannot consistently predict the “bulging” behavior of the membrane, leading to spurious estimated contacts.

### B. Results and Analysis

The contact patch estimations are shown in Figure 4 for a qualitative evaluation of contact patch geometry. The average symmetric RMSE analysis per strain state is shown in Table III. Although our proposed fusion algorithm is simple, it generalizes well across our object dataset and strain range, outperforming all of the baselines on average (Figure 7). Fusing both the tactile and proximity modalities is key to overcoming the shortcomings of each modality: the proximity modality tends to overestimate contact while tactile thresholding tends to underestimate contact.

We highlight the fusion algorithm performance for the cup, which requires segmenting two separate contacts with the base and the handle. For the medium and high strain

Low Strain (approx. 10%)	Contact Patch RMSE (mm) ↓		
Algorithm	octopus	cup	cube
Fusion	<b>1.79</b>	<b>4.96</b>	1.34
Proximity-Only	2.51	7.41	<b>1.27</b>
Tactile-Only	3.96	6.74	1.58
Mechanics Model	13.73	Fail	Fail
Medium Strain (approx. 60%)	Contact Patch RMSE (mm) ↓		
Algorithm	octopus	cup	cube
Fusion	<b>2.64</b>	<b>3.1</b>	<b>1.25</b>
Proximity-Only	4.21	8.35	3.43
Tactile-Only	4.32	6.07	1.75
Mechanics Model	9.95	10.38	6.96
High Strain (approx. 100%)	Contact Patch RMSE (mm) ↓		
Algorithm	octopus	cup	cube
Fusion	3.93	<b>2.85</b>	<b>1.57</b>
Proximity-Only	<b>2.08</b>	6.63	4.28
Tactile-Only	4.1	4.2	3.79
Mechanics Model	8.31	5.9	8.86
Overall Average for Strain Range	Contact Patch RMSE (mm) ↓		
Algorithm	low	medium	high
<b>Fusion</b>	<b>2.7</b>	<b>2.33</b>	<b>2.78</b>
Proximity-Only	3.73	5.33	4.33
Tactile-Only	4.09	4.05	4.03
Mechanics Model	13.73	9.1	7.69

TABLE III

Results from experiments across low, medium, and high strain ranges.

Average symmetric RMSE is used to calculate error.

experiments, the fusion algorithm is the only method capable of accurately distinguishing theses separate contacts.

Generally, the proximity modality overestimates contact because it cannot directly sense contact with the object. Tactile thresholding underestimates contact because there is not one threshold that can be applied across all strain ranges. Additionally, tactile thresholding struggles to distinguish two separate contacts.

There are only two specific cases where a baseline has a lower RMSE than the fused contact patch: the proximity-only estimations in the high-strain octopus contact patch and the low-strain cube patch. With both of these cases, the proximity-only baseline overestimation of the contact patch better captured the ground-truth contact patch geometry than the underestimation from the fusion algorithm. In all other cases, the fusion algorithm consistently produced the most accurate contact patch with the lowest RMSE error.

In some low strain experiments, the membrane mechanics model failed to produce an estimate of the contact patch, because the deformation is within the same magnitude as the simulated depth camera noise and thus indistinguishable. We note that the contact patch geometry estimate is dominated by the coarseness of the mesh.

## VII. DEMONSTRATIONS

### A. Complex Membrane Mechanics

1) *Varied Stiffness Membranes*: Active and passive variable stiffness materials [27]–[30] can be used for real-time optimization of robot end effector shapes and stiffnesses. Here, we demonstrate the robustness of our sensor’s multimodality and contact patch estimation to variations in the sensing membrane’s stiffness patterns (Fig. 5A). We fabricate

a varied stiffness membrane by using two base silicones, Dragonskin 10 Fast and Ecoflex 00-35 Fast (Smooth-On), with different Shore hardnesses (Figure 8). Dragonskin 10 is approximately 8x harder than Ecoflex 00-35. The stiff and soft regions are cured together to form one cohesive varied stiffness membrane.

Here, the advantage of our approach is generalization across different stiffness patterns. The proximity-only and tactile-only baseline methods would require tuning for each membrane pattern due to the variance of inflated shapes. The membrane mechanics model would require generating a new mesh that matches both the stiffness pattern and mechanical properties of the different membrane patterns. We run the medium strain cube experiment with the varied stiffness membrane and find the RMSE is 1.09 mm.

2) *Torque and Shear-Induced Wrinkling*: A highly deformable contact interface that tolerates a large range of shear and torque before slipping is key to maintaining stiction with a target object. This feature can be critical in mitigating errors in robot manipulators by reducing the requirement for precise perception and control during a task, or simplifying planning by minimizing regrasping or repositioning the object. However, these torques and shear forces can create wrinkles in the deformable membrane surface, which are very difficult to model even with computationally expensive FEM approaches. In this demonstration, we highlight the advantage of our mechanics-independent, fusion contact patch algorithm to estimate the contact patch of a cup while the membrane is wrinkled (Fig. 5B). The robot arm presses the cup into the surface of the sensor and rotates it 45°, creating wrinkles behind the cup handle and along the sides of the cup. The estimated contact patch is shown in Figure 8. In comparison with ground truth, the RMSE of the estimated contact patch is 3.49 mm. The geometry of the contact patch qualitatively matches well with ground truth, accurately capturing the side of the cup handle that is in contact with the membrane despite the wrinkles.

### B. Closed-Loop Control for Whole-Body Manipulation

We demonstrate the real-time speed and integration of the proximity and tactile fusion algorithm with a closed-loop controller to balance a stack of books (Fig. 6). The robot system, Punyo-1 [31], consists of the visuotactile and proximity sensor placed on the robot’s torso, a Kinova Jaco robot arm on each side of the sensor with soft padding, and a

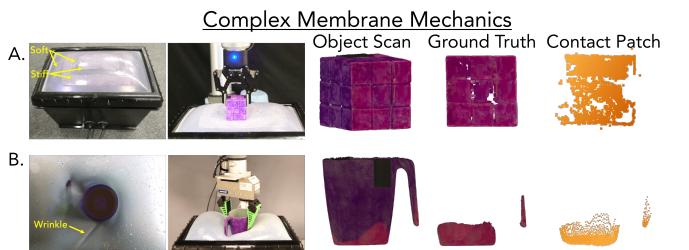


Fig. 5. We demonstrate the proposed fusion algorithm for contact patch estimation in two applications with complex membrane mechanics: A. a varied stiffness membrane and, B. membrane wrinkles, or out-of-plane deformations.

Puget computer. An empty cardboard tray is held by Punyo-1 and pressed into the sensorized torso. As each book is placed on the tray held by the robot, the tray's center of mass changes and the robot adjusts the roll angle of the tray to balance the books. The angle of the tray of books is sensed with the proximity and visuotactile fusion contact patch algorithm.

The sensor runs at 30 FPS and the contact patch is calculated for each frame. To calculate the angle of the tray, we use principal component analysis (PCA) on the contact patch. We found that PCA of the contact patch gives less noisy estimations of the tray position compared to the proximity-only modality, which has added noise from sensing the books. The estimated angle of the tray is given to the whole-body PD controller at a frequency of 29 Hz. We use two arm motion primitives to adjust the tray angle.

### C. Pose Estimation

In this demonstration, we show that our contact patches are compatible with a standard pose estimation pipeline (Fig. 7). The greater conformal contact allows for large and dense contact patch point clouds that are easy to use with a standard Iterative Closest Point (ICP) implementation from Open3D [32] to estimate pose. Exclusively using contact patches for pose estimation, rather than only the proximity modality, could be advantageous in cluttered environments where an additional segmentation step would be necessary to isolate the object of interest. We estimate the pose of a cup for 5 s as the cup simultaneously rotates about the handle and translates across the sensor surface. We assume a known model of the cup geometry. In the first frame of the trajectory, the cup is manually registered to the estimated contact patch. For the following frames that measure the rest of the trajectory, the relative transformation of the cup is estimated for each

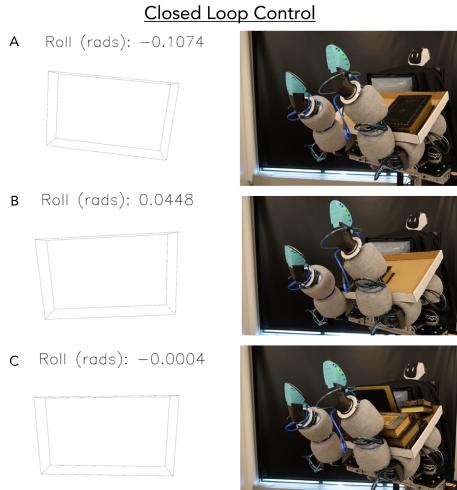


Fig. 6. The sensor is integrated on a dual arm robot platform as the chest of the robot. The sensor estimates the angle from the contact patch of the tray of books in order to balance the tray as more books are added. This task demonstrates the real-time speed and integration of the proximity and tactile fusion algorithm with a closed-loop controller. A) The tray tilts towards the right with the weight of the book. B) The tray tilts towards the left as the book slides towards the left. C) The robot holds the tray level to balance the stack of books.

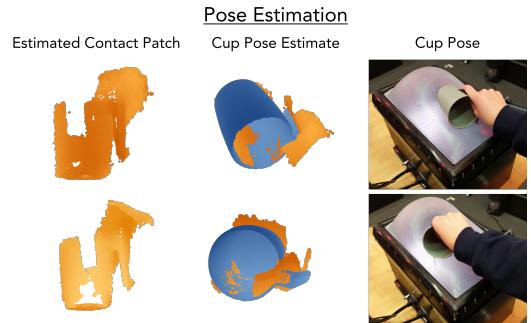


Fig. 7. The cup simultaneously rotates about the handle and translates across the membrane. The pose estimate from the contact patch qualitatively matches very well. The estimated contact patch has an atypical amount of noise near the top of the cup, which comes from the proximity estimation of the hand, due to adverse lighting conditions.

timestep. The pose estimation tends to become unstable when the cup handle is not observable. Qualitatively, the pose estimate of the cup matches well and is fairly stable across the 5 s of data. A future application of this work could be tactile SLAM for simultaneous object pose estimation and reconstruction [33], [34].

## VIII. CONCLUSION AND FUTURE WORK

In this study, we introduced a proximity and visuotactile point cloud fusion algorithm to detect contact patches with soft membranes across an unprecedented range and type of deformation. The key idea of the algorithm is that the contact patch can be identified as the set of points where the proximity point cloud and tactile point cloud intersect. Because the proximity point cloud provides localization and object geometry independent of membrane mechanics, our method is not limited by the challenge of accurately modeling the complex deformations in hyperelastic membrane materials.

We found that our proposed proximity and visuotactile fusion algorithm offered the best performance across all three strain categories. We demonstrated the use of this algorithm with three applications: complex membrane mechanics such as varied stiffness and wrinkled membranes, closed loop control for balancing a stack of books with a dual arm manipulation platform, and pose estimation.

One future direction is to explore a fingertip form factor. The current sensor size is driven by minimum sensing distances of the internal depth cameras, which can be replaced with shorter-range sensors for a significantly smaller overall package. Another future direction is to investigate using this sensor and algorithm with an actuating surface that changes stiffness and shape in real-time. A shape-changing end effector can be optimized for interactions with different objects. Our algorithm is uniquely suited to mitigate the challenges of using soft materials with variable stiffnesses and shapes.

## ACKNOWLEDGMENT

We thank the PrOps team, Aimee Goncalves, Dr. Jose Barrieros, Dr. Aykut Önal, Eric Cousineau, and Alex Alspach at TRI for helpful discussions, assistance with experiments, and hardware design advice. We thank William Yang and Greg Campbell at the Penn GRASP Lab for assistance with experiments and helpful discussions.

## REFERENCES

- [1] 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.
- [2] A. Padmanabha, F. Ebert, S. Tian, R. Calandra, C. Finn, and S. Levine, "Omnitact: A multi-directional high-resolution touch sensor," in *2020 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2020, pp. 618–624.
- [3] W. K. Do and M. Kennedy, "Densetact: Optical tactile sensor for dense shape reconstruction," in *2022 International Conference on Robotics and Automation (ICRA)*. IEEE, 2022, pp. 6188–6194.
- [4] N. F. Lepora, "Soft biomimetic optical tactile sensing with the tactip: A review," *IEEE Sensors Journal*, vol. 21, no. 19, pp. 21 131–21 143, 2021.
- [5] 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.
- [6] M. Lambeta, P.-W. Chou, S. Tian, B. Yang, B. Maloon, V. R. Most, D. Stroud, R. Santos, A. Byagowi, G. Kammerer, et al., "Digit: A novel design for a low-cost compact high-resolution tactile sensor with application to in-hand manipulation," *IEEE Robotics and Automation Letters*, vol. 5, no. 3, pp. 3838–3845, 2020.
- [7] Y. Du, G. Zhang, and M. Y. Wang, "3d contact point cloud reconstruction from vision-based tactile flow," *IEEE Robotics and Automation Letters*, vol. 7, no. 4, pp. 12 177–12 184, 2022.
- [8] Y. Zhang, Z. Kan, Y. A. Tse, Y. Yang, and M. Y. Wang, "Fingervision tactile sensor design and slip detection using convolutional lstm network," *arXiv preprint arXiv:1810.02653*, 2018.
- [9] M. Li, L. Zhang, T. Li, and Y. Jiang, "Continuous marker patterns for representing contact information in vision-based tactile sensor: Principle, algorithm, and verification," *IEEE Transactions on Instrumentation and Measurement*, vol. 71, pp. 1–12, 2022.
- [10] M. Li, T. Li, and Y. Jiang, "Marker displacement method used in vision-based tactile sensors—from 2d to 3d-a review," *IEEE Sensors Journal*, 2023.
- [11] C. Zhang, S. Cui, S. Wang, J. Hu, Y. Huangfu, and B. Zhang, "High-precision 3d reconstruction study with emphasis on refractive calibration of gelstereo-type sensors," *Sensors*, vol. 23, no. 5, p. 2675, 2023.
- [12] S. Suresh, Z. Si, J. G. Mangelson, W. Yuan, and M. Kaess, "Shapemap 3-d: Efficient shape mapping through dense touch and vision," in *2022 International Conference on Robotics and Automation (ICRA)*. IEEE, 2022, pp. 7073–7080.
- [13] M. Bauza, A. Bronars, and A. Rodriguez, "Tac2pose: Tactile object pose estimation from the first touch," *arXiv preprint arXiv:2204.11701*, 2022.
- [14] S. Suresh, Z. Si, S. Anderson, M. Kaess, and M. Mukadam, "Midas-touch: Monte-carlo inference over distributions across sliding touch," in *Conference on Robot Learning*. PMLR, 2023, pp. 319–331.
- [15] Y. Gao, S. Matsuoka, W. Wan, T. Kiyokawa, K. Koyama, and K. Harada, "In-hand pose estimation using hand-mounted rgb cameras and visuotactile sensors," *IEEE Access*, vol. 11, pp. 17 218–17 232, 2023.
- [16] 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)*. IEEE, 2017, pp. 4000–4007.
- [17] M. Oller, M. P. i Lisbona, D. Berenson, and N. Fazeli, "Manipulation via membranes: High-resolution and highly deformable tactile sensing and control," in *Conference on Robot Learning*. PMLR, 2023, pp. 1850–1859.
- [18] Y. S. Narang, K. Van Wyk, A. Mousavian, and D. Fox, "Interpreting and predicting tactile signals via a physics-based and data-driven framework," *arXiv preprint arXiv:2006.03777*, 2020.
- [19] C. Sferrazza, A. Wahlsten, C. Trueeb, and R. D'Andrea, "Ground truth force distribution for learning-based tactile sensing: A finite element approach," *IEEE Access*, vol. 7, pp. 173 438–173 449, 2019.
- [20] R. W. Ogden, G. Saccomandi, and I. Sgura, "Fitting hyperelastic models to experimental data," *Computational Mechanics*, vol. 34, no. 6, pp. 484–502, 2004.
- [21] R. Hopf, L. Bernardi, J. Menze, M. Zündel, E. Mazza, and A. Ehret, "Experimental and theoretical analyses of the age-dependent large-strain behavior of sylgard 184 (10:1) silicone elastomer," *Journal of the Mechanical Behavior of Biomedical Materials*, vol. 60, pp. 425–437, 2016. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S1751616116000758>
- [22] D. Ma, E. Donlon, S. Dong, and A. Rodriguez, "Dense tactile force estimation using gelslim and inverse fem," in *2019 International Conference on Robotics and Automation (ICRA)*. IEEE, 2019, pp. 5418–5424.
- [23] N. Kuppuswamy, A. Castro, C. Phillips-Grafflin, A. Alspach, and R. Tedrake, "Fast model-based contact patch and pose estimation for highly deformable dense-geometry tactile sensors," *IEEE Robotics and Automation Letters*, vol. 5, no. 2, pp. 1811–1818, 2019.
- [24] J. Yin, G. M. Campbell, J. Pikul, and M. Yim, "Multimodal proximity and visuotactile sensing with a selectively transmissive soft membrane," in *2022 IEEE 5th International Conference on Soft Robotics (RoboSoft)*. IEEE, 2022, pp. 802–808.
- [25] S. Brahmabhatt, C. Ham, C. C. Kemp, and J. Hays, "Contactdb: Analyzing and predicting grasp contact via thermal imaging," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2019, pp. 8709–8719.
- [26] A. Lakshminipathy, D. Bauer, and N. S. Pollard, "Contact tracing: A low cost reconstruction framework for surface contact interpolation," in *2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2021, pp. 5165–5172.
- [27] J. Pikul, S. Li, H. Bai, R. Hanlon, and I. Cohen, "Stretchable surfaces with programmable 3d texture morphing for synthetic camouflaging skins," *Science*, vol. 358, no. 6360, pp. 210–214, 2017.
- [28] R. Baines, S. K. Patiballa, B. Gorissen, K. Bertoldi, and R. Kramer-Bottiglio, "Programming 3d curves with discretely constrained cylindrical inflatables," *Advanced Materials*, p. 2300535.
- [29] G. M. Campbell, J. Yin, Y. Song, U. Gandhi, M. Yim, and J. Pikul, "Electroadhesive clutches for programmable shape morphing of soft actuators," in *2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2022, pp. 11 594–11 599.
- [30] R. Baines, S. K. Patiballa, J. Booth, L. Ramirez, T. Sipple, A. Garcia, F. Fish, and R. Kramer-Bottiglio, "Multi-environment robotic transitions through adaptive morphogenesis," *Nature*, vol. 610, no. 7931, pp. 283–289, 2022.
- [31] A. Goncalves, N. Kuppuswamy, A. Beaulieu, A. Uttamchandani, K. M. Tsui, and A. Alspach, "Punyo-1: Soft tactile-sensing upper-body robot for large object manipulation and physical human interaction," in *2022 IEEE 5th International Conference on Soft Robotics (RoboSoft)*. IEEE, 2022, pp. 844–851.
- [32] Q.-Y. Zhou, J. Park, and V. Koltun, "Open3D: A modern library for 3D data processing," *arXiv:1801.09847*, 2018.
- [33] S. Suresh, M. Bauza, K.-T. Yu, J. G. Mangelson, A. Rodriguez, and M. Kaess, "Tactile slam: Real-time inference of shape and pose from planar pushing," in *2021 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2021, pp. 11 322–11 328.
- [34] J. Zhao, M. Bauza, and E. H. Adelson, "Fingerslam: Closed-loop unknown object localization and reconstruction from visuo-tactile feedback," *arXiv preprint arXiv:2303.07997*, 2023.