Multimodal Proximity and Visuotactile Sensing with a Selectively Transmissive Soft Membrane

Jessica Yin, Gregory M. Campbell, James Pikul, and Mark Yim

***Abstract*—The most common sensing modalities found in a robot perception system are vision and touch, which together can provide global and highly localized data for manipulation. However, these sensing modalities often fail to adequately capture the behavior of target objects during the critical moments as they transition out of static, controlled contact with an end-effector to dynamic and uncontrolled motion. In this work, we present a novel multimodal visuotactile sensor that provides simultaneous visuotactile and proximity depth data. The sensor integrates an RGB camera and air pressure sensor to sense touch with an infrared time-of-flight (ToF) camera to sense proximity by leveraging a selectively transmissive soft membrane to enable the dual sensing modalities. We present the mechanical design, fabrication techniques, algorithm implementations, and evaluation of the sensor's tactile and proximity modalities. The sensor is demonstrated in three open-loop robotic tasks: approaching and contacting an object, catching, and throwing. The fusion of tactile and proximity data could be used to capture key information about a target object's transition behavior for sensor-based control in dynamic manipulation.**

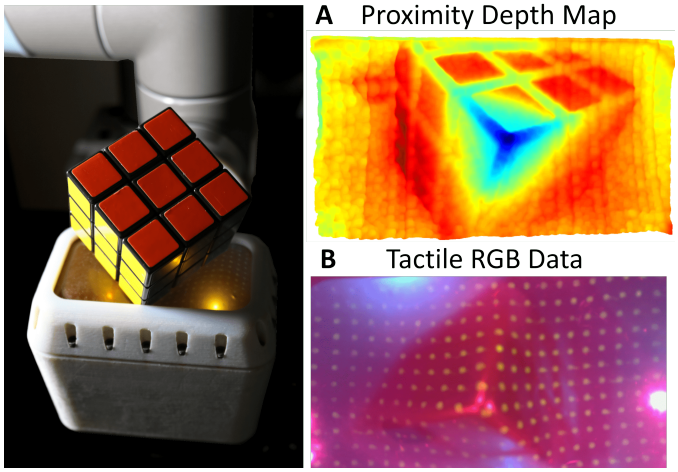

Fig. 1. A. Proximity depth map from internal depth camera. B. Image from internal RGB camera for tactile data.

I. Introduction

Approaches to perception for robot manipulation have largely mimicked the human form, focusing on the development and integration of vision sensors far from the target object and compliant tactile sensors embedded in the end effector. However, robots still struggle to achieve dexterous and dynamic manipulation capabilities comparable to humans, particularly in the case of deformable objects. This can be largely attributed to uncertainties stemming from an imperfect perception of the target object [1]. The accuracy of the object's pose estimation can make the difference between success and failure, which can be seen in "basic" tasks such as grasping, but is further amplified in dexterous and dynamic tasks that lack simple contact models and quasi-static assumptions to inform the interaction [2].

While vision sensors provide rich data about the environment and can be used to localize the target object within it, the localization estimate is not very precise – typically within a few centimeters around the object. Additionally, vision sensors are frequently occluded by robot arms as they reach towards the target or by clutter in the environment. An obvious short-term solution may be to simply add more cameras, but complete coverage of the target and workspace is not guaranteed even with multiple cameras and is not practical for real-world environments.

Tactile sensors on robot fingers and palms have been explored as a potential solution to provide more precise data about the object during contact, such as location and forces [3]. These sensors are usually designed to have mechanical compliance for increased robustness to unexpected contacts and greater functionality with the irregular or delicate geometries found in everyday objects. However, tactile sensors are only useful once the object is already in contact with the end effector, which may not be sufficient for tasks that require bringing the object in and out of contact, such as dynamic reorientation.

This points to a fundamental gap in a perception pipeline that only uses vision and touch. Closing this perception gap is necessary to create a robust perception pipeline that will allow robots to tackle more difficult manipulation tasks. A potential solution to address this gap is adding a proximity sensing modality, which can be defined as sensing within a short distance range originating from the locations of the tactile sensors [4]. Proximity sensing can provide the precise localization data that vision sensors lack and information about pre- and post-contact behavior that is difficult to predict due to complex frictional dynamics.

In this paper, we propose a novel multimodal proximity and visuotactile sensor that provides simultaneous tactile and proximity depth data. The sensor is able to detect contact over an inflated 96mm by 54mm elastomer membrane with an RGB camera (960x540) and air pressure sensor, while providing depth data with an infrared (IR) ToF camera (640x480) at a synchronous sampling rate of 30Hz. An infrared-transmissive and visibly translucent elastomer membrane, embedded with UV-phosphorescent particles, enables the simultaneous reading of visible tactile data on the membrane and IR proximity data. We introduce a sensor fusion algorithm that uses both

J. Yin, G.M. Campbell, J. Pikul, and M. Yim are with the Department of Mechanical Engineering and Applied Mechanics and the GRASP Laboratory at the University of Pennsylvania, Philadelphia, PA 19103 USA. {jessyin, gcampbel, pikul, yim} @seas.upenn.edu.

This work was supported by the National Science Foundation Graduate Research Fellowship Program under Grant No. 2020295381 and NSF #1935294.

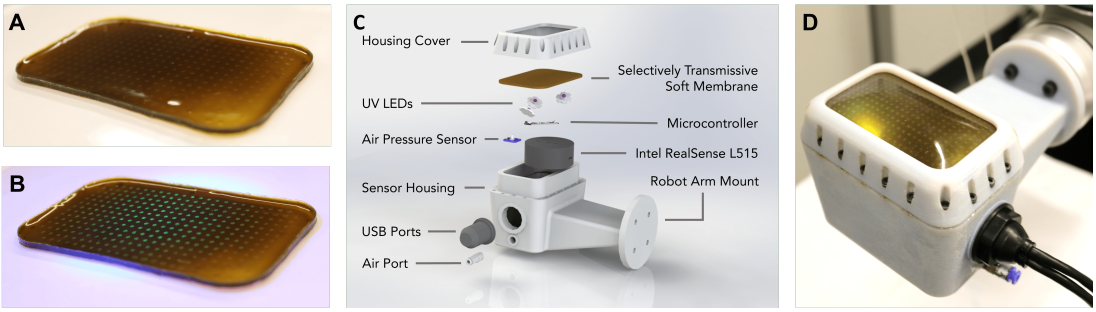

Fig. 2. A. Soft membrane in ambient room light. B. Soft membrane with UV-phosphorescent dot grid pattern activated by 365nm UV light. C. Overview of sensor system. D. Sensor mounted on UR10 robot arm.

the RGB image and depth image to correct for the effect of the embedded particles in the depth image, and evaluate the depth data up to 100mm from the sensing surface. The sensor is integrated into an end-effector and mounted on a UR10 (Universal Robots) robot arm and demonstrated with the following open-loop tasks: approach and contact, catching, and throwing.

II. Design and Fabrication

A. Selectively Transmissive Soft Membrane

We designed the soft membrane for selective transmission to achieve the following: (1) allow the infrared light (860nm) emitted by the time-of-flight camera to pass through, (2) block most of the visible light (400nm-700nm) from the external environment, and (3) enable the activation of the phosphorescent, light-emitting particles (∼500nm) on the inner surface from internal UV LEDs (365nm). Blocking external visible light enhances the visual contrast with the green-colored phosphorescent particles (Figure 2A, 2B), facilitating the facile application of off-the-shelf OpenCV algorithms for tracking [5]. Additionally, the membrane is designed to be physically resilient for repeated use in contact-rich interactions, while providing a highly compliant contact surface. The thickness of the membrane can be decreased for greater infrared light transmissivity, but at the cost of reduced physical robustness and opacity to external light.

We fabricate the membrane in layers; each silicone elastomer layer fully cures prior to pouring the next layer. The first layer consists of Ecoflex 00-30 mixed by hand with a dye solution (Epolight 7276B; Epolin, dissolved in chloroform, 5.325g/L concentration) in a 15:1 (mL) elastomer to dye solution ratio. The dye solution is visibly opaque and infrared transmissive. We pour 8.25g of the dyed elastomer into an Ease-Release-coated laser cut mold and place it in a vacuum degassing chamber for 10 minutes. Then, the mold is placed on a hot plate and heat-cured at 100°C for 10 minutes.

The next layer consists of the UV-phosphorescent particles in a dot grid pattern. The UV-phosphorescent particles are made of Cu:ZnS (copper doped zinc sulfide, 35 microns; Technoglow). We mixed 0.2g of Cu:ZnS with 2g of Ecoflex 00-30 by hand. We laser cut a 0.508mm stencil made of clear PVC to the shape of the membrane and desired dot grid pattern (1mm diameter, 4mm uniform spacing, 328 total dots). We press the stencil onto the membrane to remove air bubbles and the Cu:ZnS elastomer mixture is spread onto the surface with a q-tip. The membrane cures on a hot plate at 100°C for 10 minutes.

The final layer evens out the protrusions from the dot grid layer and leaves a slightly matte finish to reduce specular reflections from the infrared and UV lights. We mix Ecoflex 00-30 with NOVOCS Matte silicone solvent (Smooth-On) in a 3:6 (g) solvent to elastomer ratio and degas for 10 minutes. Finally, we pour 4g of the mixture onto the membrane and cure at 100°C on a hot plate for 10 minutes.

B. Internal Electronics

We chose to use the Intel Realsense L515 because it provides integrated RGB and ToF depth cameras, as well as the ability to adjust the ToF laser power to bring the minimum sensing range to approximately 50mm for short range sensing. The field of view (FOV) of the RGB camera is 70° by 43°and the FOV of the depth camera is 70° by 55°. Because the FOVs don't exactly align, the active sensing area only consists of the overlapping regions of both FOVs. The cameras output data through the same USB-C port, which is connected to an airtight USB-A 3.0 port that goes through the sensor housing. The internal air pressure sensor samples data at 30Hz and is connected to a microcontroller. We inflate the membrane to a gauge pressure of 0.02PSI to reduce the specular reflection of the internal UV and IR lights.

We soldered three UV LEDs (365nm) onto aluminum heat sinks and connected them in series with 50mΩ resistance. The LED circuit is connected to the direct USB power output pin on the microcontroller, which provides 2.1A. The microcontroller is connected to the USB-A port that goes through the sensor housing. An overview of the internal components of the sensor system is shown in Figure 2C.

III. Sensor Characterization

A. Proximity Depth Sensing

In this section, we characterize the proximity depth sensing. A test stand with discrete slots from 10mm-100mm in 10mm increments mounts a flat plane parallel to the sensing surface. Four sheets of white printer paper (92 brightness) cover the flat plane and encompass the entire FOV of the depth sensing. We apply the dot correction algorithm and average 20 consecutive frames for evaluation of the data. The average depth pixel value of the depth data has an R^2 = 0.725 fit with the ground truth distance.

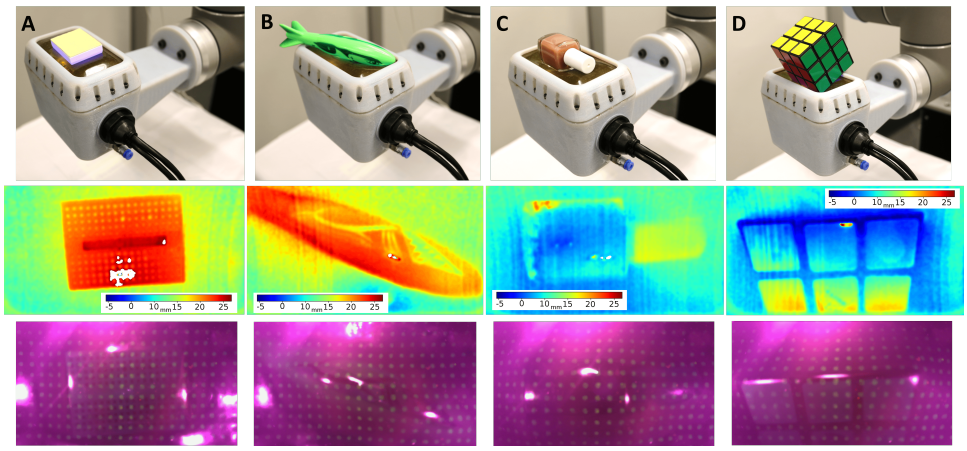

Fig. 3. Top row: test object, middle row: corresponding proximity depth map, bottom row: corresponding tactile RGB image with phosphorescent dots for membrane tracking. A. Solderless breadboard. B. Shark torpedo. C. Nail polish. D. Rubik's cube.

Figure 3 shows depth maps of objects placed on the surface of the sensor. We compared the depth maps to the significant dimensions of each object and found an average overall error of 4.3%. The sensor showed poorer performance on curved surfaces, with the average error of 5.5% and much better on edges, with an average error of 2%.

B. Tactile Sensing

Tactile sensing is achieved by measuring the change in the internal air pressure and by tracking the motion of the dots on the internal membrane surface. The dots are detected in the RGB image with the simple blob detector and tracked with the Lucas-Kanade optical flow algorithms from OpenCV. The simple blob detector finds the center coordinates of each dot in each frame, and then the optical flow calculates the distance between its initial position and current position. To detect contact, the total flow velocity summed from all the dots act as a proxy for the magnitude of membrane deformation, and therefore total contact force. An RGB image of an uncontacted and inflated membrane initializes the optical flow and subsequent frames are compared to the uncontacted state. The air pressure sensor uses gauge pressure for contact detection.

Measuring both the internal air pressure and flow velocity for binary contact detection extends the range of contact that can be detected. The internal air pressure is more sensitive to contact and can detect forces below 100g, which are not sufficient to create an appreciable change in flow velocity. The flow velocity is particularly useful for detecting tangential forces and lateral motions of the object along the sensing surface, which may not produce significant changes in the air pressure. The sensitivity of the flow velocity contact detection can be tuned to detect different ranges of forces by changing the window size of the optical flow algorithm.

IV. Demonstrations

We mounted the sensor to a UR10 robot arm and demonstrated tasks where it could be beneficial to use both proximity and tactile sensing modalities. Each task can be separated into "pre-contact", "during contact", and "post-contact" stages that excite both sensing modalities. Furthermore, accurate perception of object behavior in all of these stages is critical for successful manipulation.

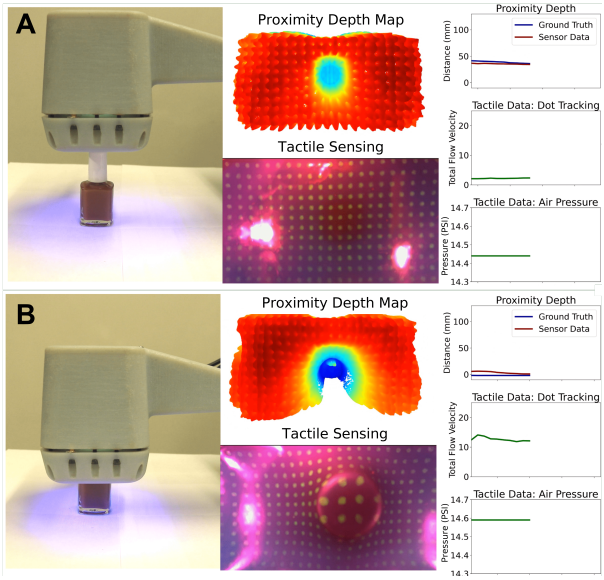

Fig. 4. Proximity depth and tactile sensing data: A. Before the sensor makes contact with the nail polish. B. After the sensor makes contact with the nail polish.

A. Approach and Contact

The robot performs an approach and contact task with a bottle of nail polish placed on a table. The pose of the end effector from the robot arm provides the ground truth for the distance accuracy of the sensor. Data from the same approach and contact trial are shown in Figure 4.

The sensor's initial pose is oriented towards the nail polish, 40mm above the object (Figure 4A). The proximity depth map shows the top of the nail polish clearly in the center, while the tactile RGB image shows zero deformation of the sensing surface. The proximity plot on the left estimates the nail polish is 40mm away from the sensing surface, which is in excellent agreement with the ground truth. The tactile data plots on the right also show data that corroborates with the no-contact

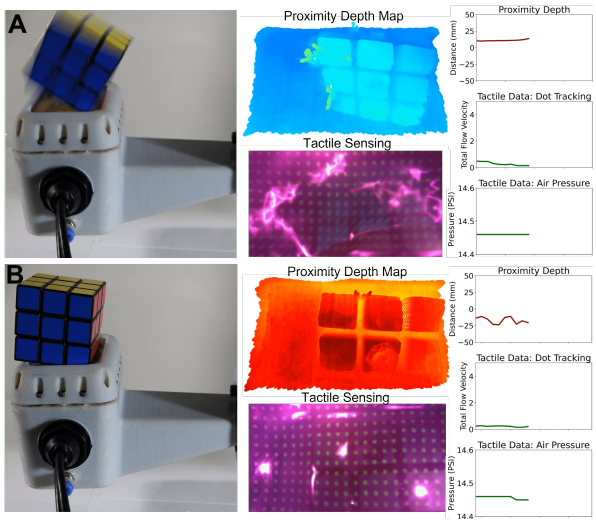

Fig. 5. Proximity depth and tactile sensing data: A. Before the Rubik's cube makes contact with the sensor. B. After the Rubik's cube makes contact.

state: the dot tracking flow velocity is close to zero, and the air pressure has not changed from the initial air pressure.

The robot arm then moves the sensor towards the bottle at 50mm/s until it makes contact and protrudes 10mm into the sensing surface (Figure 4B). The proximity depth map shows the top of the bottle protruding into the surface and outlines the body of the bottle, while the tactile RGB image shows the circular top of the bottle producing a deformation of the sensing surface. The plots on the right show that the proximity estimation is in good agreement with the ground truth and that the nail polish is in contact with the sensing surface. The tactile data plots show a significant increase in dot tracking flow velocity from the membrane deformation, as well as a 0.15PSI increase in internal air pressure.

B. Catching

To demonstrate catching, the arm-mounted sensor faces the ceiling and a Rubik's cube is dropped onto the sensor from a height of 80mm. Figure 5 consists of data from the same catching trial, prior to and after contact. All x-axes on the rightmost plots are time-synchronized.

Figure 5A shows the cube just before making contact with the sensor. The proximity depth map shows the angular orientation and the square features of the cube, while the RGB tactile image shows no deformation. The plots on the right show that the cube is 7mm above the sensing surface, and the air pressure and dot tracking flow velocity show little to no change from initial no-contact conditions.

Figure 5B displays data after the cube settled onto the sensing surface. The proximity depth map and plot shows about half of the cube protruding 22mm into the membrane. The tactile RGB image shows some deformation of the membrane surface, with a small increase in dot tracking flow velocity.

The proximity depth data senses the cube 13 frames before it makes contact with the membrane. Both the proximity depth data and tactile sensing data match well qualitatively with a video recording of the experiment.

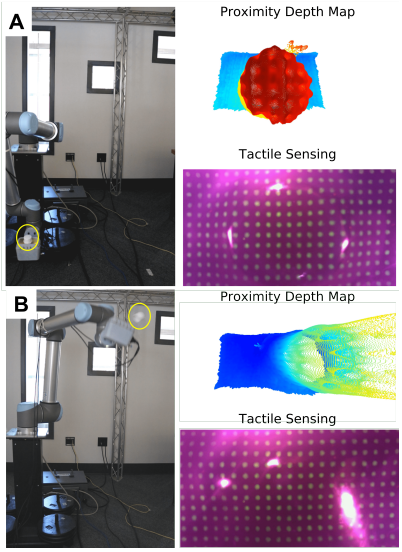

Fig. 6. Proximity depth and tactile sensing data, with the target object circled in yellow: A. Prior to throwing the hex head cap. B. After throwing the hex head cap.

C. Throwing

In this demonstration, the UR10 arm throws a 46mm diameter PVC hex head cap off the surface of the sensor (Figure 6). The final speed of the end effector is approximately 1.5m/s. Until the hex head cap is thrown, it remains in contact with the sensor, although we observe some lateral rocking during the wind-up trajectory in both the tactile and proximity data. Capturing this type of object behavior could be very useful in predicting object trajectories after a throw. The tactile and proximity data both show loss of contact at the end of the throw. After the hex head cap is thrown, the sensor captures 6 frames of the hex head cap's initial projectile motion.

V. Conclusions and Future Work

In this study, we introduced a novel multimodal proximity depth and visuotactile sensor enabled by a selectively transmissive elastomer membrane. We presented the design and fabrication techniques for each component of the sensor, and we evaluated the proximity depth data across a distance range of 10mm-100mm. Both the binary contact detection and proximity depth modalities were tested with an object dataset consisting of nail polish, Rubik's cube, breadboard, shark torpedo, and PVC hex head cap.

We integrated the sensor into an end-effector to mount on a UR10 robot arm and demonstrated it in three open-loop tasks where the mixed modality of the sensor could provide an advantage. The demonstrations and quality of the data show potential for the application of this sensor to capture target object behavior before, during, and after contact in dynamic and dexterous manipulation tasks.

Moving forward, we plan to work on the fusion of tactile and proximity data for contact patch and force estimation. Due to the nonlinear mechanics introduced by the use of a highly deformable elastomeric membrane, it is challenging to precisely relate force, deformation, and geometry. With force estimations and expanded tactile sensing capabilities, the

sensor could be used to develop dynamic and dexterous control policies using proximity and tactile data.

VI. References

[1] Aude Billard and Danica Kragic. "Trends and challenges in robot manipulation". In: *Science* 364.6446 (2019).

[2] Fabio Ruggiero, Vincenzo Lippiello, and Bruno Siciliano. "Nonprehensile dynamic manipulation: A survey". In: *IEEE Robotics and Automation Letters* 3.3 (2018), pp. 1711–1718.

[3] Qiang Li, Oliver Kroemer, Zhe Su, et al. "A review of tactile information: Perception and action through touch". In: *IEEE Transactions on Robotics* 36.6 (2020), pp. 1619–1634.

[4] Stefan Escaida Navarro, Stephan Mühlbacher-Karrer, Hosam Alagi, et al. "Proximity Perception in Human-centered Robotics: A Survey on Sensing Systems and Applications". In: *IEEE Transactions on Robotics* (2021).

[5] Gary Bradski and Adrian Kaehler. "OpenCV". In: *Dr. Dobb's journal of software tools* 3 (2000).